Journal of Experimental Psychology

ARTHUR W. MELTON, Editor
Department of Psychology, University of Michigan
Ann Arbor, Michigan

DAVID A. GRANT, Associate Editor
University of Wisconsin

DELOS D. WICKENS, Associate Editor
Ohio State University

Consulting Editors

E. JAMES ARCHER, *University of Wisconsin*
JUDSON S. BROWN, *State University of Iowa*
CLETUS J. BURKE, *Indiana University*
PAUL M. FITTS, *Ohio State University*
FREDERICK C. FRICK, *Massachusetts Institute of Technology*
FRANK A. GELDARD, *University of Virginia*
JAMES J. GIBSON, *Cornell University*
CLARENCE H. GRAHAM, *Columbia University*
HAROLD W. HAKE, *University of Illinois*
LLOYD G. HUMPHREYS, *University of Illinois*
ARTHUR L. IRION, *Tulane University*
HOWARD H. KENDLER, *New York University*
DONALD B. LINDSLEY, *University of California, Los Angeles*
KENNETH MACCORQUODALE, *University of Minnesota*
QUINN MCNEMAR, *Stanford University*
L. STARLING REID, *University of Virginia*
KENNETH W. SPENCE, *State University of Iowa*

CONTENTS

Stimulus and Response Generalization: Tests of a Model Relating Generalization to Distance in Psychological Space: R. N. SHEPARD 509
Acquisition and Change of a Concept Attitude as a Function of Consistency of Reinforcement: R. J. RHINE AND B. A. SILUN 524
The Relation of Visual Acuity to Convergence and Accommodation: E. TULVING 530
Reproductive Interference Following "Appropriate" and "Inappropriate" Warm-Up Activities: L. E. THUNE .. 535
The Relation of Eye Movements, Body Motility, and External Stimuli to Dream Content: W. DEMENT AND E. A. WOLPERT 543
Associative Interference in Verbal Paired-Associate Learning: N. F. BESCH AND W. F. REYNOLDS 554
Effects of Stress and Anxiety on Performance of a Complex Verbal-Coding Task: L. T. KATCHMAR, S. ROSS, AND T. G. ANDREWS 559
Mediated Stimulus Equivalence and Distinctiveness in Human Conditioning: G. R. GRICE AND J. D. DAVIS 565
Serial Verbal Learning Under Two Conditions of Hunger Motivation: R. G. LERNER, I. SINGER, AND H. C. TRIANDIS 572
Achievement Motivation and Goal-Setting Behavior on a Learning Task: D. H. KAUSLER AND E. P. TRAPP 575
Effect of Knowledge of Conditioning Upon Eyelid Conditioning: W. R. MCALLISTER AND D. E. MCALLISTER ...579
Habit Reversal as a Function of Schedule of Reinforcement and Drive Strength: H. H. KENDLER AND R. LACHMAN 584
Comparing Reversal and Nonreversal Shifts in Concept Formation With Partial Reinforcement Controlled: M. HARROW AND G. B. FRIEDMAN 592
Summation of Manifest Anxiety and Muscular Tension: D. R. MEYER AND M. E. NOBLE 599
Variation of Temporal Intervals Among Critical Events in Five Studies of Knowledge of Results: E. A. BILODEAU AND I. MCD. BILODEAU 603
Temporal Effects of Conditioned Fear on the Eyelid Reflex: K. W. SPENCE AND W. N. RUNQUIST 613
Binaural Supplementation of the Auditory Intensive Threshold: W. A. SHAW 617
Progressive Ambiguity in the Attainment of Concepts on the Wisconsin Card Sorting Test: I. GORMEZANO AND D. A. GRANT 621
Sequential Decision Making: Wald's Model and Estimates of Parameters: G. M. BECKER 628
The Relationship Between Discriminability and Generalization: A Re-Evaluation: H. I. KALISH 637

American Psychological Association

Vol. 55, No. 6 June 1958

ARTHUR C. HOFFMAN, Managing Editor
HELEN ORR, Circulation Manager
SADIE J. DOYLE, Editorial Assistant

The JOURNAL OF EXPERIMENTAL PSYCHOLOGY is published monthly, two volumes per year, by the American Psychological Association, Inc., at Prince and Lemon Sts., Lancaster, Pa. The subscription rate per volume is $8.00, or $16.00 annually. Single copies are $1.50. Subscriptions, orders, address changes, and business communications should be addressed to the American Psychological Association, Inc., 1333 Sixteenth St. N.W., Washington 6, D. C.

This JOURNAL publishes original experimental investigations which are intended to contribute toward the development of psychology as an experimental science. Studies with normal human subjects are favored over studies involving abnormal or animal subjects, except when the latter are specifically oriented toward the extension of behavior theory. Experimental psychometric studies and studies in applied experimental psychology or engineering psychology may be accepted if they have broad implications for experimental and theoretical psychology.

Normally, articles exceeding 20 printed pages cannot be accepted. Within this limit, the piecemeal experiment-by-experiment reporting of psychological research is discouraged, and the reporting of the data and conclusions of substantial segments of research efforts is encouraged. Specifically, an integrated series of studies accomplished simultaneously (e.g., most doctoral dissertations) *must* be presented in a single article. Special provision is made for the publication of Supplementary Reports and Replication Reports in articles of not more than 1.5 printed pages (see Editor's Note, this JOURNAL, 1957, **53**, 1–2).

Address all articles submitted for publication to the Editor: Arthur W. Melton, Department of Psychology, University of Michigan, Ann Arbor, Michigan. **Manuscripts *must* adhere to the conventions concerning reference citations, preparation of tables and figures, manuscript format, etc., as described in the *Publication Manual of the American Psychological Association*.** When in doubt about practices in this Journal, authors should examine a recent issue.

Articles are published in the order of their receipt, except in rare circumstances. Authors are supplied with 50 free offprints without covers. All of the cost of an author's alterations in galley proof is charged to the author. Priority in publication is given to articles whose authors assume the full cost of publication. In 1958 the publication cost is $20 per printed page. Authors of priority publications receive no free offprints.

To save printing costs supplementary material to articles in this Journal are deposited with the American Documentation Institute. Copies may be obtained by ordering the Document Number given in the footnote reference to ADI. Orders should be sent to ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C. Orders *must* be accompanied by checks or money orders payable to the Chief, Photoduplication Service, Library of Congress.

Entered as second-class matter, February 6, 1937, at the post office at Lancaster, Pa., under the Act of March 3, 1879. Acceptance for mailing at the special rate of postage provided for in paragraph (d-2), Section 34.40, P. L. & R. of 1948, authorized September 11, 1947.

Send address changes to 1333 Sixteenth St. N.W., Washington 6, D. C. Address changes must reach the Subscription Office by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

© 1958 by the American Psychological Association, Inc.

Journal of
Experimental Psychology

VOL. 55, NO. 6 JUNE, 1958

STIMULUS AND RESPONSE GENERALIZATION: TESTS OF A MODEL RELATING GENERALIZATION TO DISTANCE IN PSYCHOLOGICAL SPACE[1]

ROGER N. SHEPARD

Harvard University

If, with a given organism, a stimulus, S_a, leads to a response, R_a, and this sequence, $S_a \rightarrow R_a$, is followed by reinforcement, two effects subsequently appear. First, the conditional probability of R_a, given a different stimulus S_b, is increased if S_b is similar to S_a (stimulus generalization). Second, the conditional probability of a different response R_b, given S_a, is increased if R_b is similar to R_a (response generalization). Thus, neither a stimulus nor a response is an individual unit that can be operated upon in isolation. Rather, the various stimuli and responses are linked together so that the consequences of reinforcement are necessarily transmitted from the original stimulus and response to other neighboring or similar stimuli and responses which were not manifest on the occasion of reinforcement.

Although the phenomena of generalization are governed by interstimulus and interresponse similarities, it is usually presumed that the precise form of the relation between generalization and similarity can be determined only if there is available an independent measure of similarity (**23,** p. 361; **25,** p. 20; **32,** p. 755). But this strategy immediately encounters a dilemma: (*a*) Similarity could be objectively specified by physical measurement or by the summation of JND's. In either case, however, the gross form of the relation between similarity and generalization would vary from one set of stimuli or responses to another (**26,** p. 100; **32,** pp. 577–579), and this relation may not even be monotonic (**4, 8, 12**). (*b*) Similarity could be directly estimated by judgmental procedures. But to found an investigation of the elementary and pervasive phenomena of generalization upon the more complex processes

[1] This paper is based upon the experimental portion of a Ph.D. dissertation submitted to the Graduate School of Yale University and upon subsequent investigations pursued on a National Academy of Sciences—National Research Council Postdoctoral Research Associateship at the Naval Research Laboratory. The author is particularly indebted to C. I. Hovland, R. P. Abelson, and B. S. Rosner for their many valuable suggestions. In addition, G. A. Miller, W. D. Garvey, and J. G. Holland supplied a number of helpful comments; Patricia A. Rund and Jean B. Henson provided technical assistance; and W. J. McGuire generously made his unpublished data available for analyses undertaken here.

of introspection and verbal communication seems backward. (*c*) Finally, similarity could itself be determined from the data of previous experiments on generalization. But this is clearly circular.

Fortunately it is possible to proceed without an independent measure of similarity and, thus, to circumvent the entire dilemma. The decisive step consists in replacing the notion of similarity with that of distance; for the concept of distance (unlike that of similarity) has a rigorous interpretation in the form of a set of metric axioms. The strategy, then, becomes one of discovering what function will convert data from experiments on generalization into distances (i.e., numbers satisfying the metric axioms). This strategy avoids the circularity of Alternative *c* of the dilemma; for it will always be a strictly empirical question whether the numbers entailed by the assumption of any particular function will or will not satisfy the metric axioms.

Because the metric axioms express intuitively compelling properties of the ordinary notion of distance, the dependence of the present strategy upon these axioms should not be viewed as a disadvantage. Rather, the disadvantage accompanies any presumptive measure of distance which has not been shown to satisfy these axioms; for the sense in which such a measure can be called a distance remains in doubt. This is clear from an examination of these axioms which, for present purposes, are reducible to the following two:

I. The distance between any point (i.e., stimulus or response) and itself is zero:

$$D_{ii} = 0. \tag{1}$$

II. The length of one side of a triangle formed by any three points S_i, S_j, and S_k cannot exceed the sum of the lengths of the other two sides:

$$D_{ik} \leqslant D_{ji} + D_{jk}, \tag{2}$$

and the length of that one side equals the sum of the other two sides,

$$D_{ik} = D_{ji} + D_{jk}, \tag{3}$$

if and only if the triangle is collapsed so that the point S_j falls on the straight line connecting S_i and S_k. Relations 2 and 3 will be referred to as the triangle inequality and the additivity property, respectively.

The procedure for testing whether generalization is related to distance according to some particular function, f, is carried out as follows: First, a set of presumptive distances is computed by applying the inverse function, f^{-1}, to the generalization data. The resulting presumptive distances are then "forced" to satisfy the metric axioms and, thus, are converted into "true" distances. The "forcing" is accomplished (in a procedure analagous to factor analysis) by extracting a set of coordinates giving the position of each point in a metric space. The true distances are then computed from these coordinates using the generalized Pythagorean formula and can be shown to satisfy the metric axioms. Finally, application of the postulated function, f, to these true distances generates a reconstructed set of data which is subject to direct comparison with the original data. If the reconstructed and the original data are consistent, it may be concluded that the "forcing" is unnecessary and, by implication, that f is indeed the correct function. The mathematical details of this procedure are given in a theoretical paper (**28**).

A stochastic model proposed by Shepard (**28**) is based on the assumptions (*a*) that generalization is an exponential decay function of psychological distance, and (*b*) that the two processes of stimulus and response generalization are independent. In testing these assumptions, the present experiments are concerned exclusively with multiple stimulus-response or paired-associate situations. Such situations illustrate particularly clearly the descriptive economy of the spatial characterization of the stimuli and responses.

Before proceeding to these experiments, however, it should be noted that the above discussion has assumed that each stimulus and each

TABLE 1

GENERALIZATION DATA FOR NINE STIMULI VARYING IN SIZE

Stimulus	Conditional Probabilities*									Weights W_i^S	Coordinates X_i^S
	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9		
S_1	.804	.118	.041	.013	.009	.005	.004	.001	.004	1.647	−4.332
S_2	.068	.584	.201	.090	.019	.005	.015	.004	.014	1.196	−2.321
S_3	.040	.170	.378	.237	.092	.020	.043	.006	.014	.788	−1.371
S_4	.034	.049	.175	.492	.109	.054	.063	.006	.018	.932	− .608
S_5	.038	.032	.066	.285	.267	.170	.104	.014	.024	.454	.107
S_6	.018	.011	.024	.100	.106	.410	.252	.050	.030	.715	1.024
S_7	.014	.017	.019	.034	.026	.236	.401	.216	.036	.924	1.538
S_8	.003	.011	.007	.008	.006	.090	.153	.528	.195	.891	2.479
S_9	.010	.010	.010	.003	.001	.024	.039	.246	.657	1.139	3.485

* The symbols along the left refer to the stimuli presented and the symbols across the top refer to the stimuli taken by *S*s to have been presented. Thus S_4 was mistaken for S_3 with a probability of .175, and S_4 was correctly identified with a probability of .492. The stimuli are designated in accordance with size so that S_1 denotes the smallest circle, S_2 the next largest, etc.

response has the same intrinsic status as every other, and that it is only the relation (i.e., the distance) between them that controls generalization. Actually, the stimuli or responses may differ widely in "familiarity" (**24, 28**), "insistence" (**5,** p. 40; **14,** p. 108), "intensity" (**11**), or "preference" (**2, 15, 16**). In these cases a weight can be estimated for each stimulus or response which accounts for any asymmetric tendency to perceive one stimulus or emit one response more frequently than some other (**28**). If the weights are required, this will be attested by the gain in accuracy realized when the weights are included (with the distances) in the reconstruction of the original data.

REANALYSIS OF MCGUIRE'S DATA

In order to test the hypothesis that stimulus generalization is an exponential decay function of psychological distance, the responses should be so distinctive as to contribute negligibly to the occurrence of generalization errors, and the stimuli should vary over a restricted range of a single physical dimension (**28**). Fortunately, suitable data are already available from an experiment carried out by McGuire in another connection (**18**).

In McGuire's experiment 10 undergraduate students served as *S*s. Nine black circles were presented. Their diameters varied in equal steps from .37 to .91 cm. On each of 720 successive presentations one of the nine circles was exposed for 4 sec. and *S* responded by speaking a number between "one" and "nine" during the first 2 sec. of this exposure. During the second half the number assigned to the presented circle was exposed along with that circle, and this reinforcement by correction served to build up the correct association between the responses and the stimuli. Each *S* learned one of two assignments of the responses to the stimuli. Both assignments were arranged so that there was no correlation between the size of the circle and the value of its assigned response number.

For present purposes the data from McGuire's experiment were retabulated in a 9×9 matrix giving for every pair of stimuli, S_i and S_k, the conditional probability with which S_i led to the response assigned to S_k, averaged over all trials and *S*s (Table 1). Under the assumption that the responses were almost never confused, an entry P_{ik}^S in this matrix can be considered as an estimate of the probability that S_i was taken (through a stimulus generalization) to be S_k (**28**).

Since the stimuli were generated by a relatively small variation of a single physical parameter, they should fall along a straight line in "psychological space" (**28**), and the presumptive interstimulus distances should be additive. The computational procedures set forth in the theoretical paper were therefore applied to the data of Table 1 to yield, for each stimulus S_i, a weight, W_i^S, and a coordinate, X_i^S, giving the location of that stimulus on the common straight line, i.e., on the single axis of the one-dimensional psychological space. These 18 quantities are also presented in Table 1. In addition to these, a single constant, $C^S = .008$, was calculated, in accordance with the model, to account for the essentially random responses which occur during

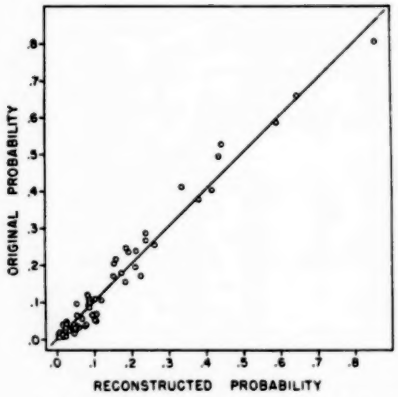

FIG. 1. Probabilities of stimulus generalizations between circles varying in size as a function of probabilities reconstructed from the coordinates and weights for these stimuli.

the early learning trials. The 9×9 matrix of true interstimulus distances (i.e., distances satisfying the metric axioms including additivity) can be directly obtained from the nine spatial coordinates by the one-dimensional distance formula

$$D_{ik}^{S} = |X_i^{S} - X_k^{S}| \qquad (4)$$

(where the vertical strokes indicate that the coordinate difference is always taken as positive).

Using the equations of the model, a reconstruction of the 81 original probabilities of Table 1 was attempted on the basis of the nine weights and nine coordinates in Table 1, together with the constant C^{S}. The probabilities of Table 1 are plotted against these reconstructed probabilities in Fig. 1. With respect to the variance of the original probabilities, the proportion accounted for in the reconstruction is .973. The proportion accounted for when the weights are not included in the reconstruction is .957. Thus the weights appear to be relatively unimportant as compared with the distances. That almost 96% of the variance of the original matrix (with 72 *df*) can be accounted for in terms of just 10 numbers (with 9 *df*), points to the extreme redundancy of the original data. The accuracy of the reconstruction, moreover, lends considerable support to the basic assumption under test; namely, that generalization is an exponential decay function of psychological distance.

EXPERIMENT I

The data from McGuire's experiment were particularly useful because the stimuli varied along only one psychological dimension and thus permitted a test of the postulated exponential decay function under the stringent condition that the distances be additive. However, those data cannot be used to test the applicability of the postulated function to situations in which the distances involve variations along more than one psychological dimension. In view of Attneave's conclusion that a metric treatment may not be feasible in such cases (**1**), Exp. I was designed to test the postulated function under this less stringent but more general condition.

Method

Subjects.—The *S*s were 36 students with normal color vision according to the Ishihara Test (**13**).

Stimuli.—Nine colored circles of uniform size were presented. These colors, as specified by the system of Munsell (**20**), were of a constant red hue, but varied in brightness (value) and saturation (chroma), as shown in Table 2. Care was taken to insure that responses were learned to the colors per se rather than to incidental markings. First, squares cut from Munsell color sheets were mounted behind ¾-in. circular holes cut in the centers of 5×7-in. white cards. Second, 13 such cards were prepared for each of the nine colors. Thus, during learning no card was seen more than twice by any one *S*.

Procedure.—For each *S* the stimuli were exposed in a random sequence of 200 successive

TABLE 2

MUNSELL DESIGNATION FOR NINE STIMULI VARYING IN COLOR

Stimulus	Hue	Value	Chroma
S_1	5R	7	4
S_2	5R	6	6
S_3	5R	6	10
S_4	5R	5	4
S_5	5R	5	8
S_6	5R	5	12
S_7	5R	4	6
S_8	5R	4	10
S_9	5R	3	4

presentations. The responses, again the spoken numbers "one" through "nine," were discriminatively attached to the stimuli through the method of correction. The stimulus for each presentation was removed and the next exposed after *S* had responded and the correction had been given. In order to counterbalance any response generalization, each *S* was trained under a different assignment of the responses to the stimuli. Using a permutation procedure developed for the purpose (**28**), the assignments were selected so that, over all 36 *S*s, every pair of responses was assigned to each pair of stimuli just once.

Results

The data were tabulated in a 9×9 matrix which, if it can be assumed that the responses were almost never confused, can be taken to give the probabilities of stimulus generalization $P_{ik}{}^S$ (Table 3). Exactly the same procedures as those applied to McGuire's data were used to calculate, from Table 3, a weight $W_i{}^S$, a coordinate $X_{1i}{}^S$ (giving the projection for S_i on the first dimension), and a coordinate, $X_{2i}{}^S$ (giving the projection for S_i on the second dimension) for each stimulus, S_i. These 27 quantities are also presented in Table 3. Again, a constant, $C^S = .020$, was estimated to account for the early random responses. The 9×9 matrix of true distances can be calculated from the 18 spatial coordinates by the two-dimensional Pythagorean formula

$$D_{ik}{}^S = \sqrt{(X_{1i}{}^S - X_{1k}{}^S)^2 + (X_{2i}{}^S - X_{2k}{}^S)^2}. \quad (5)$$

The results in Table 3 can be evaluated in terms of internal consistency, reliability, and conformity with the scaling results of other investigators using entirely different methods.

First, the distances computed from the coordinates by means of Equation 5 necessarily satisfy the metric axioms (including the triangle inequality but, of course, excluding the property of additivity). However, it remains to be determined whether the original data were violated in the process of satisfying these axioms. In order to rule out this possibility, a reconstruction of the 81 original probabilities of Table 3 was attempted on the basis of the 27 weights and coordinates, together with the constant C^S. The original data are compared with the results of the reconstruction in Fig. 2. The proportion of the original variance accounted for is .993. The proportion accounted for when the weights are not included is .988. Once again the stimulus weights apparently play a negligible role as compared with the distances. That approximately 99%

TABLE 3

GENERALIZATION DATA FOR NINE STIMULI VARYING IN COLOR

Stimulus	Conditional Probabilities									Weights $W_i{}^S$	Coordinates	
	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9		$X_{1i}{}^S$	$X_{2i}{}^S$
S_1	.678	.148	.054	.030	.025	.020	.016	.011	.016	1.051	2.427	−2.410
S_2	.167	.544	.066	.077	.053	.015	.045	.018	.015	1.002	1.048	−1.672
S_3	.060	.070	.615	.015	.107	.067	.022	.030	.014	0.933	2.287	0.316
S_4	.015	.104	.016	.542	.057	.005	.163	.032	.065	1.022	−1.187	−1.291
S_5	.037	.068	.120	.057	.460	.075	.057	.099	.030	0.772	0.397	0.642
S_6	.027	.029	.053	.015	.036	.715	.015	.095	.014	1.055	1.533	3.124
S_7	.011	.033	.015	.145	.049	.016	.533	.052	.145	0.957	−1.794	−0.263
S_8	.016	.027	.031	.046	.069	.096	.053	.628	.034	0.932	−0.761	2.280
S_9	.005	.016	.011	.068	.020	.021	.061	.018	.780	1.205	−3.949	−0.726

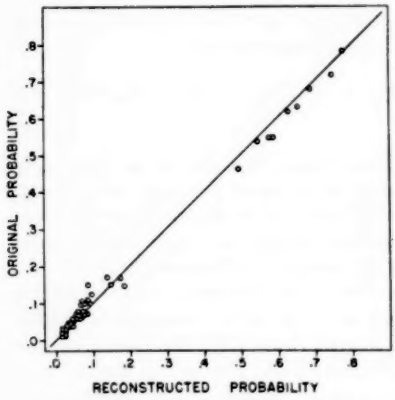

FIG. 2. Probabilities of stimulus generalizations between circles varying in color as a function of probabilities reconstructed from the coordinates and weights for these stimuli.

of the original variance (with 72 *df*) can be attributed to the 18 coordinates and constant C^S (with a total of 17 *df*), however, indicates that the spatial representation is consistent with the experimental data. That just two coordinate-axes are required for the stimulus characterization can be verified by similar considerations. Since the maximum variance accounted for by just one dimension (together with the weights) is only 64.8% of the original, a set of coordinates along a second dimension is indicated. A third set would be of little significance, however, since only .7% of the variance remains to be accounted for after the second dimension has been added.

The reliability of the results was assessed by dividing the 36 *S*s, at random, into two subgroups of 18 each. The coordinates and weights were then independently calculated for each of these groups. The product-moment r between the coordinates for the two groups (with 16 *df*) was .96. Since this correlation is statistically significant at beyond the .01 level, it may be concluded that the process of representing the stimuli in psychological space is repeatable. The r between the weights for the two groups (with 8 *df*), however, was only .48. Since this is not significant at the .05 level, the hypothesis that the weights do not contribute anything beyond the spatial characterization cannot be rejected for these stimuli.

Finally, the present spatial characterization can be compared with the following five multidimensional scaling solutions, previously obtained for the same stimulus colors, using judgmental procedures:

Munsell I: The 1929 standardization of the coordinates for the Munsell system (**20, 31**).

Munsell II: The 1943 revision of the coordinates for the Munsell system (**21**).

Torgerson: The coordinates obtained by Torgerson using his multidimensional method of complete triads (**30,** Ch. 11).

Messick I: The coordinates obtained by Messick in a replication of Torgerson's procedure (**19**).

Messick II: The coordinates obtained by Messick using his multidimensional method of successive intervals (**19**).

The locations of the stimuli in psychological space according to the coordinates of these five, as well as the present scaling solution, are shown in Fig. 3.[2]

An objective measure of the agreement between any two of these scaling solutions is the proportion of their coordinate variance held in common after the two solutions have been brought into alignment by a similarity transformation (**3,** p. 259). (Such a transformation is required because most of the scaling solutions are determined only up to an arbitrary translation, rotation, and expansion of the reference axes.) If the reference origin for each solution has

[2] For purposes of comparison, all of the solutions were rotated to correspond with the Munsell representation in terms of brightness (value) and saturation (chroma). Since Messick did not include S_4, only the eight stimuli common to all six solutions are considered. Also, in order to relate Munsell's coordinates to those of the other solutions, some assumption had to be made regarding the relative psychological distances entailed by one step of value as opposed to one step of chroma. On the basis of Nickerson's conclusions (**22**), two chroma units were taken to be psychologically equivalent to one unit of value.

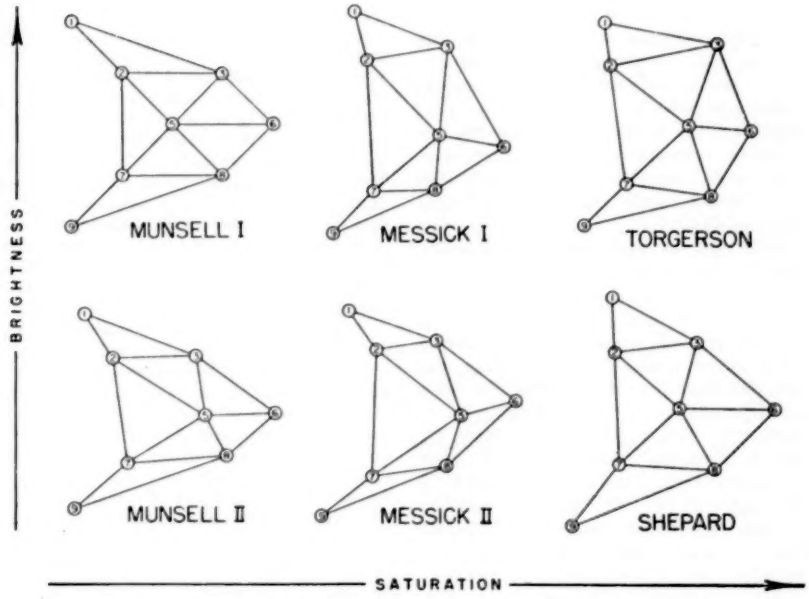

FIG. 3. Spatial configurations for eight stimuli varying in color according to six different scaling solutions. Each circle gives the position in psychological space of a single stimulus according to its associated stimulus coordinates.

been translated to the centroid of the configuration of stimulus-points, the formula for the common proportion of coordinate variance after transformation to best fit, as previously derived (**27**), is

$$\frac{[\Sigma(X_{1i}^S)(Y_{1i}^S) + \Sigma(X_{2i}^S)(Y_{2i}^S)]^2 + [\Sigma(X_{1i}^S)(Y_{2i}^S) - \Sigma(X_{2i}^S)(Y_{1i}^S)]^2}{[\Sigma(X_{1i}^S)^2 + \Sigma(X_{2i}^S)^2][\Sigma(Y_{1i}^S)^2 + \Sigma(Y_{2i}^S)^2]}. \quad (6)$$

Here the X coordinates are for one solution, the Y coordinates are for another, and the index of summation, i, ranges over all stimuli. The measures of agreement as computed by this formula are presented in Table 4. In view of the uniformity of this table, there does not appear to be sufficient

TABLE 4

PROPORTION OF COORDINATE VARIANCE COMMON TO EACH PAIR OF SCALING SOLUTIONS

	Mean*	Messick II	Messick I	Torgerson	Munsell II	Munsell I
Shepard	.972	.97	.97	.99	.97	.96
Messick II	.965	—	.97	.96	.97	.93
Messick I	.965		—	.98	.94	.92
Torgerson	.968			—	.95	.95
Munsell II	.956				—	.97
Munsell I	.938					—

* The two solutions Munsell I and II are not independent. Therefore, Munsell I was not included in the computation of the mean proportion for the other solutions.

evidence for rejecting the hypothesis that each solution is an approximation to the same underlying configuration. This is particularly striking in the present procedure which yields results that share, on the average, 97% of the coordinate variance with the other solutions, even though these other solutions were based upon radically different types of data.

Although, in evaluating the generalization model, principal reliance should be placed on the internal tests, it is interesting to note that the obtained probabilities of stimulus generalization could have been predicted from the previous multidimensional stimulus characterizations based upon judgments of similarity. Thus if two parameters, C^8 and a distance multiplier, had been estimated from the data of Table 3 (leaving 70 *df*), Torgerson's scaling solution could have been used to predict 97% of the variance in that table.

Experiment II

The McGuire experiment and Exp. I were suitable for studying stimulus generalization but not response generalization. For, whereas the stimulus similarities were high, the response similarities were quite low. In order to investigate response generalization, therefore, the difficulty must be shifted from the stimulus to the response side of the paired-associate task. Accordingly, Exp. II is the converse of McGuire's experiment, i.e., the stimuli are highly discriminable but the responses vary over a restricted range of a single physical dimension.

Method

Subjects.—The 36 *S*s upon which the analysis is based included both employees of the Naval Research Laboratory and enlisted men assigned there. The *S*s varied widely in intelligence and motivation and it was found necessary to eliminate five of the original 41 *S*s who, after 300 trials, were still responding at chance level. (Subsequent interviews with these *S*s revealed that they had misconstrued the instructions.)

Responses.—On each stimulus presentation *S* was required to insert an electric probe into a narrow rectangular slot in the top of the box schematically illustrated in Fig. 4. Nine ¼-in. metal contacts were arranged in a row along the bottom of the 2¼-in. slot. The depth of the slot and the angle of illumination were adjusted so that *S* could not see the individual contacts. The relay circuitry governing the presentation of stimuli and recording of responses was set up so that (*a*) as soon as the probe touched any of the nine contacts, the trial was terminated, (*b*) the probe had to touch a tenth contact, external to the slot, in order to initiate the next trial, and (*c*) double responses could not occur.

Procedure.—On each of 300 successive trials, one of the nine symbols (a, b, c, 1, 2, 3, X, Y, Z) of the vertical 3 × 3 array shown in Fig. 4 was selected at random and illuminated from behind. (However, all nine symbols were visible at all times.) For each *S* there was a different assignment of the responses to these stimuli such that, over all 36 *S*s, every pair of responses was assigned to every pair of stimuli just once. Thus any residual stimulus generalization could be counterbalanced in the manner previously described (**28**). The association between the stimuli and their assigned responses was built up, for each *S*, by stimulus correction, i.e., the symbol corresponding to the contact touched was illuminated more brightly than the symbol chosen as the actual stimulus on that trial. Thus, whereas an incorrect response caused a second, brighter light to flash on, a correct response simply caused the light already on to flash brighter. (The relation between this

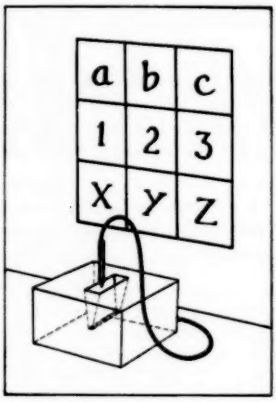

Fig. 4. Apparatus for studying response generalization, and the associated stimulus-display.

TABLE 5

GENERALIZATION DATA FOR NINE RESPONSES VARYING IN POSITION

Response	Conditional Probabilities*									Weights W_i^R	Coordinates X_i^R
	R_1	R_2	R_3	R_4	R_5	R_6	R_7	R_8	R_9		
R_1	.834	.033	.027	.025	.022	.009	.007	.010	.034	2.119	−4.447
R_2	.129	.294	.204	.108	.083	.066	.054	.035	.026	0.656	−2.126
R_3	.057	.145	.367	.177	.100	.076	.048	.015	.014	1.033	−1.327
R_4	.045	.067	.217	.346	.160	.071	.050	.026	.017	1.014	−0.597
R_5	.036	.046	.092	.213	.305	.191	.068	.030	.019	0.925	0.091
R_6	.035	.031	.071	.099	.171	.322	.156	.075	.038	0.832	0.761
R_7	.026	.045	.086	.093	.135	.187	.283	.108	.038	0.687	1.437
R_8	.041	.058	.078	.099	.109	.141	.196	.209	.067	0.434	2.006
R_9	.036	.014	.027	.022	.021	.035	.031	.046	.770	1.345	4.201

* The symbols along the left refer to the responses presumably intended and the symbols across the top refer to the responses actually made by *Ss*. The responses are designated on the basis of position. Thus R_1 denotes the response contact that is at the greatest distance from *S*, etc.

stimulus correction technique and the response correction technique used in Exp. I will be considered in the discussion.)

Results

The data were tabulated in a 9×9 matrix giving for every pair of responses, R_i and R_k, the conditional probability with which the stimulus assigned to R_i led to R_k (Table 5). Under the assumption that the stimuli were almost never confused, an entry P_{ik}^R in this matrix can be considered as an estimate of the probability that R_k was made (through a response generalization) in place of R_i.

Since the responses varied over a small range of a single physical dimension, they should fall along a straight line in psychological space (**28**), and the presumptive interresponse distances should be additive. The computational procedures were therefore applied to the conditional probabilities in Table 5 to obtain, for each response R_i, a weight W_i^R, and a coordinate X_i^R, giving the location of that response along the common axis in psychological space. These 18 quantities are also presented in Table 5, and the estimated locations of the nine response points as specified by the spatial coordinates, X_i^R, are plotted in Fig. 5. Again, the interresponse distances satisfying the metric axioms are given in terms of the coordinates by the one-dimensional formula

$$D_{ik}^R = |X_i^R - X_k^R|. \qquad (7)$$

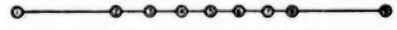

FIG. 5. Spatial configuration for nine responses varying in position. Each circle gives the position in psychological space of a single response according to its associated coordinate.

As may be seen in Fig. 5, the seven intermediate responses are spaced quite evenly. The two extreme responses, however, are situated at a considerably greater distance from the others. This is an unavoidable consequence of the design of the apparatus (as is clear in Fig. 4). In order to touch either end contact, *S* had simply to slide the probe down the partition at the end of the slot. This distinctive aspect of these responses served to enhance their discriminability or psychological distance from the others. Unfortunately, the estimation of C^R is useful only to the extent that the variation among the distances is not too great (**28**). In the present case this problem is best

handled by separately estimating two constants: one, $C_{(int)}{}^{R} = .149$, for the seven intermediate responses and one, $C_{(ext)}{}^{R} = .042$, for the two extreme responses.

A reconstruction of the 81 probabilities of Table 5 was attempted on the basis of the 18 weights and coordinates, plus the two constants. The results are shown in Fig. 6. With respect to the variance of the original probabilities, the proportion accounted for by the reconstruction is .992. The proportion accounted for when the weights are not included in the reconstruction is .921. The accuracy of the reconstruction again supports the postulation of an exponential decay relation between generalization and psychological distance.

The reliability of the response characterization was assessed by dividing the 36 *Ss*, at random, into two subgroups of 18 each. The product-moment *r*'s between the results for the two groups, calculated independently, were, for the coordinates, .99, and, for the weights, .94 (both highly significant with 8 *df*). Since the response weights are reliably contained in the data, they are plotted against the corresponding stimulus coordinates in Fig. 7. The pattern obtained indicates a tendency to favor the ends and the center of the one-dimensional response array.

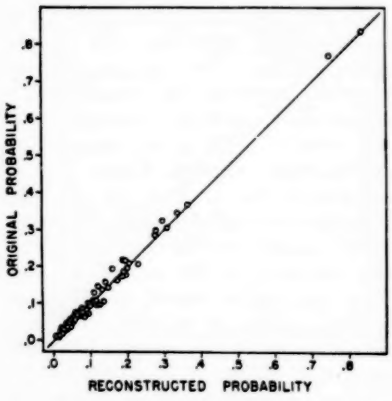

FIG. 6. Probabilities of response generalization between responses varying in position as a function of probabilities reconstructed from the coordinates and weights for these responses.

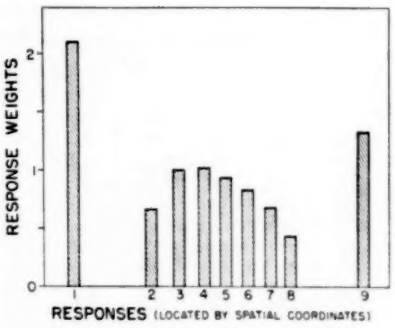

FIG. 7. Response weights as a function of the corresponding response coordinates for nine responses varying in position.

EXPERIMENT III

In each of the preceding experiments there was an appreciable degree of generalization, either among the stimuli or among the responses, but never among both at once. Thus, although these experiments support the assumption of the exponential decay function, they do not provide an efficient test of the assumption that the stimulus and response generalization processes are mutually independent. In order to test this assumption, the interstimulus distances and the interresponse distances must simultaneously be made quite small in a single experiment. In Exp. III, therefore, the color-stimuli of Exp. I are combined with the probe responses of Exp. II in a new paired-associate task. If the independence assumption is correct, a prediction of the experimental results of Exp. III should be possible on the basis of the coordinates and weights already calculated separately for the stimuli and the responses.

Method

Subjects.—The *S*s included both employees of the Naval Research Laboratory and naval enlisted men. As in Exp. II, only the 36 *S*s with better than chance performances at the end of 300 trials were included in the analysis. Also, owing to limitations in the supply of available personnel, half of the *S*s had already served in Exp. II. However, four months separated the two experiments, and the similarity of the results for the old and new *S*s suggested that transfer effects were negligible.

Stimuli and responses.—Because the two end responses of Exp. II necessitated the estimation of a separate constant, $C_{(ext)}{}^{S}$, it was decided to eliminate these responses from consideration in Exp. III. In order to do this, seven of the color stimuli from Exp. I were assigned to the seven intermediate responses of Exp. II. Colors S_1 and S_9, however, were replaced by two new and highly distinctive stimuli (a black triangle and a crosshatched square) which were assigned to the end responses. In this way generalization errors associated with the two end responses and the two new stimuli were reduced to a minimum. The analysis, then, treats the data for the seven remaining stimuli and responses only. A justification for this procedure is provided, in part, by the "constant-ratio rule" enunciated by Clarke (**7**).

Procedure.—Each *S* was required to respond during each of 300 successive stimulus presentations. In contrast to the procedures of the previous experiments the same assignment of the responses to the stimuli prevailed for all *S*s in this experiment. The particular assignment arbitrarily chosen is indicated in Fig. 8. Also, in order to avoid the asymmetries of the stimulus or the response-correction techniques, a straight correct-incorrect method of reinforcement was instituted. If *S* responded correctly on a given trial, a $\frac{1}{2}$-sec. 500-cycle tone signaled this fact. If *S* responded incorrectly, the tone was omitted.

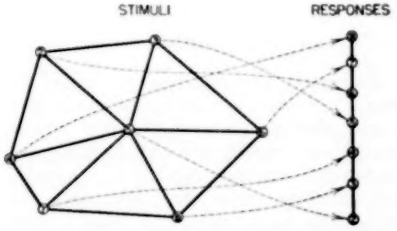

FIG. 8. Spatial configurations for seven stimuli varying in color and for seven responses varying in position. The stimulus-response assignment used in Exp. III is indicated by the broken arrows.

Results

The data were tabulated in a 7×7 matrix giving for any color stimulus, S_i, the probability with which it leads to any of the seven intermediate responses, R_k. This matrix was then compared with a predicted matrix, calculated on the basis of the coordinates and weights for the seven colors and seven probe positions.

Since Exp. I was run for only 200 trials, it was necessary, before making the prediction, to multiply the interstimulus distances by a constant factor to render them equivalent to distances based upon 300-trial data. In order to estimate this factor, the average interstimulus distance was separately evaluated on the basis of the cumulative data up to each 20-trial division of the 200 trials. A smooth curve was then drawn through these points and extrapolated to determine what the final average for the interstimulus distance would have been if the experiment had been continued for 300 trials. The ratio of the average distance at 300 to that at 200 trials provides the desired factor of 1.136.

The predicted matrix was calculated by substituting the following 37 quantities into the central equation of the model: the 14 stimulus coordinates (multiplied by 1.136) and the 7 stimulus weights given in Table 3, the 7 response coordinates and the 7 response weights given in Table 5, and the two constants $C^S = .020$ and $C^R = .149$. The prediction was complete in the sense that no parameters were estimated from the empirically obtained matrix. The obtained data are plotted against the predicted data in Fig. 9. With respect to the

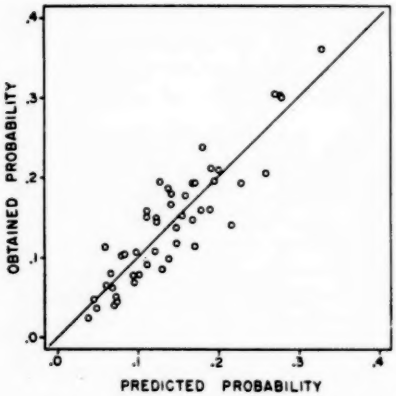

FIG. 9. Conditional probabilities with which each color stimulus led to each position response. The empirically obtained probabilities are plotted as a function of the probabilities predicted from the coordinates and weights for the stimuli and responses.

variance of the obtained probabilities, the proportion accounted for in the prediction is .82 (with 42 *df*).

This proportion (though statistically significant) is lower than the proportions of variance reconstructed in the preceding experiments. However, it remains to be determined whether the variance not accounted for in the prediction is to be attributed to a violation of the independence assumption or to an intrinsic unreliability of the data. In order to make this determination, the 36 *S*s were divided at random into two subgroups of 18 each. The average variance accounted for, by prediction from the model, in these two subgroups is .76. The average accounted for, by prediction from one subgroup to the other, is .68. Thus it appears that prediction from the model is at least as accurate as prediction from one experiment to a replication of that experiment. This suggests that the model predicts as precisely as the error of measurement will permit. The independence assumption cannot, therefore, be rejected on the basis of these data. The unexpected variability of the data precludes a sensitive test, however. (That the prediction from model to data is somewhat more accurate than the prediction from data to data raises no statistical problems since the prediction from the model was based upon 72 *S*s whereas the prediction from the data was based upon only 18 *S*s. In general, the correlation between a stable estimate and an unstable estimate is greater than the correlation between two unstable estimates of the same quantity.)

Finally, the predictive power of the independence postulate is illustrated by the fact that the present experiment is only one out of 7! experiments for which predictions could have been made. This is because only one of the 7! possible assignments of the responses to the stimuli was chosen for this experiment. Thus, with the 37 initial quantities characterizing the stimuli and responses, a total of $7 \times 7 \times 7!$ or 246,960 conditional probabilities could have been predicted. Indeed, the power is perhaps even greater than this since 18 of the initial quantities (i.e., the weights) would contribute only slightly to the accuracy of these predictions.

DISCUSSION

Before relating the present results to those obtained in previous investigations, certain assumptions upon which the present results are based should be examined. It was explicitly assumed, for example, that the responses in McGuire's experiment and Exp. I and the stimuli in Exp. II were almost never confused. But this is surely justified by the fact that, whereas the constants C^S and $C_{(\text{ext})^R}$ in these experiments were quite small, the proportion of the original

variance which could be reconstructed was in all cases greater than .97 (**28**).

An implicit assumption was also made regarding the correction techniques of reinforcement employed in McGuire's experiment and Exp. I and II. The model was constructed on the supposition that nonreinforcement consists merely in indicating the occurrence of an incorrect response (e.g., as in Exp. III by the absence of a tone) without providing further information about the size or direction of the error. Except for a few investigations, such as those by Thorndike (**29**), this method of reinforcement has been employed only rarely in paired-associate tasks. Usually the method of correction or, more accurately, the method of response correction is used. In McGuire's experiment and Exp. I it was assumed that the practice of indicating the correct response when it is not made does not affect the occurrence of stimulus confusions, though it may diminish the probability of response confusions. In fact this method was used because it was desired to keep response confusions to a minimum. Since in Exp. II, on the other hand, it was desired to minimize stimulus confusions, a stimulus correction method was employed. This procedure is the converse of the response-correction method in that the correction specifies, not the response assigned to the presented stimulus, but the stimulus assigned to the elicited response. Just as it was assumed that supplying response corrections does not alter the pattern of stimulus confusion, it is assumed here that providing stimulus corrections does not influence the pattern of response confusion. That these assumptions are not greatly in error is suggested by the accuracy of the prediction from these experiments to the results of Exp. III.

The existence of these two alternative correction procedures raises the question of whether there is any connection between (*a*) the statement sometimes made that, whereas stimulus generalization is an important factor in the learning of paired associates, response generalization does not play a comparable role, and (*b*) the fact that the method of response correction is almost universally employed. Indeed, it is perhaps for similar reasons that doubt has been expressed about the primary status of response generalization as a phenomenon of conditioning (**10**, p. 617). For, in the conditioning situation differential reinforcement is set up just so that response generalization will be eliminated. That is, incorrect responses, whenever made, go unreinforced. Such differential training, however, does not ordinarily take place on the stimulus side since incorrect stimuli are never presented. The analogue on the stimulus side, of the differential training on the response side, would be to present a variety of stimuli similar to the correct stimulus but reinforce only when the response is made to the correct stimulus. In this way stimulus generalization could be eliminated also.

According to the present model, stimulus and response generalization are equally primary and essentially analogous phenomena. If one wishes to establish a unique connection between a particular stimulus S_a and response R_a, one must both (*a*) extinguish incorrect responses to S_a and (*b*) extinguish R_a to incorrect stimuli. During the early trials, of course, any stimulus may be present when R_a occurs and any response may occur in the presence of S_a. Thus, from this standpoint, response generalization necessarily operates in exactly those situations with respect to which its operation has been expressly doubted (**17**, p. 327), i.e., in the absence of relevant prior learning. The interesting fact about response generalization, however, is not that any response may initially occur but that during differential training it is the more psychologically distant responses that extinguish first.

The present experiments taken together, then, appear to lend substantial support to the assumption that the probability of a generalization error is an exponential decay function of the distance between the two stimuli or responses involved. The objections

raised by Bush and Mosteller (**6**) to the specification of such a generalization function, moreover, are excluded by defining distance in terms of a set of metric axioms rather than in terms of some physical measure of dissimilarity.

Furthermore, the evidence regarding the possibility of representing a set of stimuli as points in a psychological space is more encouraging than might have been anticipated from Attneave's discussion (**1**). Indeed, that stimuli can be scaled solely on the basis of overt errors during learning suggests that the present method may implement the study of the perceptual behavior of young children and infrahuman animals. For such purposes, moreover, the applicability of this scaling procedure to stimuli differing along an unknown number of dimensions sets it in contrast with those related procedures, like the "equal discriminability" scaling technique of Garner and Hake (**9**), which can be conveniently applied only to stimuli that are uniquely ordered along a single one-dimensional line.

With respect to the weights, the present evidence indicates that these are not of practical importance in accounting for errors of generalization on the stimulus side. However, the relative homogeneity of the stimuli selected may have precluded the observation of the type of perceptual asymmetry for which the stimulus weights were introduced.

Summary

Three experiments and one by McGuire (**18**) on the learning of paired associates are used to test the basic assumptions of a stochastic model relating generalization to psychological distance, where distance is defined in terms of a set of metric axioms. McGuire's experiment (using, as stimuli, circles varying in size) and Exp. I (using, as stimuli, circles varying in brightness and saturation) support the assumption that stimulus generalization is an exponential decay function of psychological distance between the stimuli. Experiment II (employing, as responses, the various possible placements of a probe in a narrow slot) supports the assumption that response generalization is an exponential decay function of psychological distance between the responses. And, although it did not provide conclusive evidence, Exp. III (using the stimuli of Exp. I and the responses of Exp. II) was consistent with the assumption that the processes of stimulus and response generalization are mutually independent.

In addition, it is demonstrated that the stimuli or responses can be represented as points in a psychological space of one or more dimensions solely on the basis of the overt errors made during learning. Such a representation is economical in the sense that it permits a reconstruction of the empirical data from a small set of stimulus or response coordinates in that space. Once the stimuli and the responses have been given a spatial characterization, moreover, the postulated independence of the processes of stimulus and response generalization provides for prediction to a wide range of new paired-associate experiments using these stimuli and responses.

References

1. Attneave, F. Dimensions of similarity. *Amer. J. Psychol.*, 1950, **63,** 516–556.
2. Barber, J. L., & Garner, W. R. The effect of scale numbering on scale-reading accuracy and speed. *J. exp. Psychol.*, 1951, **41,** 298–309.
3. Birkhoff, G., & MacLane, S. *A survey of modern algebra.* New York: Macmillan, 1953.
4. Blackwell, H. R., & Schlosberg, H. Octave generalization, pitch discrimination, and loudness thresholds in the white rat. *J. exp. Psychol.*, 1943, **33,** 407–419.
5. Bruner, J. S., Goodnow, J. J., & Austin, G. A. *A study of thinking.* New York: Wiley, 1956.
6. Bush, R. R., & Mosteller, F. A model for stimulus generalization and discrimination. *Psychol. Rev.*, 1951, **58,** 413–423.
7. Clarke, F. R. Constant-ratio rule for confusion matrices in speech communication. *J. acoust. Soc. Amer.*, 1957, **29,** 715–720.
8. Duncan, C. P. Development of response generalization gradients. *J. exp. Psychol.*, 1955, **50,** 26–30.
9. Garner, W. R., & Hake, H. W. The amount of information in absolute judgments. *Psychol. Rev.*, 1951, **58,** 446–459.
10. Hovland, C. I. Human learning and retention. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951.

11. Hull, C. L. Stimulus intensity dynamism (*V*) and stimulus generalization. *Psychol. Rev.*, 1949, **56**, 67–76.
12. Humphreys, L. G. Generalization as a function of method of reinforcements. *J. exp. Psychol.*, 1939, **25**, 361–372.
13. Ishihara, S. *Tests for colour-blindness.* (5th ed.) Tokyo: Kanehara, 1925.
14. Katz, D. *The world of colour.* London: Kegan Paul, 1935.
15. Kaufman, E. L., Lord, M. W., Reese, T. W., & Volkmann, J. The discrimination of visual number. *Amer. J. Psychol.*, 1949, **62**, 498–525.
16. Lincoln, R. S., & Alexander, L. T. Preferred patterns of motor and verbal responses. *J. exp. Psychol.*, 1955, **50**, 106–112.
17. McGeoch, J. A., & Irion, A. L. *The psychology of human learning.* New York: Longmans, Green, 1952.
18. McGuire, W. J. A multi-process model for paired-associates learning. Unpublished Ph.D. dissertation, Yale Univer., 1954.
19. Messick, S. J. An empirical evaluation of multidimensional successive intervals. *Psychometrika*, 1956, **21**, 367–375.
20. *Munsell Book of Color.* (Abridged ed.) Baltimore: Munsell Color Co., Inc., 1945.
21. Newhall, S. M., Nickerson, D., & Judd, D. B. Final report of the O.S.A. subcommittee on the spacing of the Munsell colors. *J. opt. Soc. Amer.*, 1943, **33**, 385–418.
22. Nickerson, D. The specification of color tolerances. *Textile Res.*, 1936, **6**, 505–514.
23. Osgood, C. E. *Method and theory in experimental psychology.* New York: Oxford Univer. Press, 1953.
24. Pillsbury, W. B. A study in apperception. *Amer. J. Psychol.*, 1897, **8**, 315–393.
25. Plotkin, L. Stimulus generalization in Morse code learning. *Arch. Psychol. N. Y.*, 1943, **40**, 287.
26. Rothkopf, E. Z. A measure of stimulus similarity and errors in some paired-associate learning tasks. *J. exp. Psychol.*, 1957, **53**, 94–101.
27. Shepard, R. N. Stimulus and response generalization during paired-associates learning. Unpublished Ph.D. dissertation, Yale Univer., 1955.
28. Shepard, R. N. Stimulus and response generalization: A stochastic model relating generalization to distance in psychological space. *Psychometrika*, 1957, **22**, 325–345.
29. Thorndike, E. L. An experimental study of rewards. *Teachers Coll. Contr. Educ.*, 1933, No. 580.
30. Torgerson, W. S. *Theory and methods of scaling.* New York: Wiley, 1958.
31. Tyler, J. E., & Hardy, A. C. An analysis of the original Munsell color system. *J. opt. Soc. Amer.*, 1940, **30**, 587–590.
32. Woodworth, R. S., & Schlosberg, H. *Experimental psychology.* New York: Holt, 1955.

(Received May 8, 1957)

Journal of Experimental Psychology
Vol. 55, No. 6, 1958

ACQUISITION AND CHANGE OF A CONCEPT ATTITUDE AS A FUNCTION OF CONSISTENCY OF REINFORCEMENT

RAMON J. RHINE[1] AND BETSY A. SILUN

University of Massachusetts

A great deal of effort has been expended on the study of existing attitudes and on investigations of attitude change, but there is only a relatively small amount of controlled experimentation dealing with the details of attitude development. Eisman's (**2**) research stands almost alone in this area although studies such as those by Horowitz (**9**) and by Clark and Clark (**1**) are pertinent to the understanding of attitude growth.

Perhaps more research would be obtained if attitude was expressed in a meaningful way that would make it possible to apply laboratory techniques to attitude learning. As at least a first approximation toward attaining this goal, an attitude may be thought of as a concept with an evaluative dimension. This point of view is related to the thinking of Osgood and Tannebaum (**13**).

Typically an *S* in a concept-formation experiment must learn that, of a series of experiences, only those which have certain common properties are instances of a concept which he is set to discover. He is often presented with a series of objects which differ in any of several characteristics, e.g., size, shape, and color. After many experiences, he may learn that *E* will tell him he has chosen the correct object only when he picks those that are both tall and black. Since no other properties will do, he soon learns that he can choose correctly on the basis of the concept *tallness plus blackness*.

Similarly, in real life, he may learn that objects with the dimensions "human, dark skin, kinky hair, and thick lips" are instances of a concept which is labeled *Negro*. As such, his concept of Negro is not an attitude. However, when an additional evaluative dimension like "goodness" or "badness" is also acquired as part of the concept, there is a concept attitude. In the present example, the concept attitude might be based on the dimensions "human, dark skin, kinky hair, thick lips, and *bad*."

Two features of this concept-attitude approach should be mentioned. First, when attitudes are interpreted within a concept-formation frame of reference, existing explanations of concept learning are applicable. Second, a method for studying attitudes under controlled conditions follows from the concept-attitude view.

The experiment which follows employs this method to explore concept-attitude development and change as a function of consistency of reinforcement and stimulus intensity. Consistency of reinforcement is studied by observing how concept-attitude development and change proceed when a high proportion of "bad" properties are attributed to a fictional people in comparison to situations

[1] Now at the System Development Corporation, Santa Monica, Calif. This research was supported in part by a Faculty Research Grant from the University of Massachusetts. The data for this study were collected by the junior author under the direction of the senior author in partial fulfillment of the requirements for Honors in Psychology.

where relatively few "bad" characteristics and relatively many "good" characteristics are attributed to the people. This use of different reinforcement schedules is similar to procedures employed in studies of verbal probability learning (**7, 8, 10**). The present experiment also investigates concept-attitude development and change under conditions where the number of favorable or unfavorable properties attributed to the people are constant, but the language used to describe them is relatively weak or strong. This aspect of the experiment emphasizes differing stimulus intensities.

Method

Subjects

A preliminary study employed 100 undergraduates, and another 80 *Ss* were used in the main experiment. All *Ss* were drawn from the introductory psychology course at the University of Massachusetts.

Materials

In order to obtain concept-formation materials with an empirical value dimension, the 100 *Ss* were given 206 trait names which they were asked to rate as "very desirable" in a person or as "fairly desirable," "fairly undesirable," or "very undesirable." Some example traits are: active, boastful, gifted, homosexual, lazy, refined, and rude. Only traits rated either as "desirable" (positive traits) or "undesirable" (negative traits) with 90% or greater agreement were retained for the experiment. Each of these traits was printed in large letters on a 4 × 6-in. index card. Of the negative traits with 90% or greater agreement, those most often rated "very undesirable" were separated from those most often rated "fairly undesirable."

Procedure

The 80 experimental *Ss* were instructed that a little-known people had been visited by anthropologists, missionaries, physicians, and others, and that these visitors were in high agreement about a number of personality traits of these people. The *Ss* were told that *E* wanted to determine if they could predict the traits attributed to the people. They were also informed that *E* would tell them after each prediction if the anthropologists and others had felt that the trait was characteristic of the people.

Concept-attitude development.—After these instructions, *Ss* were shown 50 traits, one at a time, and they wrote their predictions for each one separately. The 50 traits included 25 positive and 25 negative words placed in random sequence.

Consistency of reinforcement was varied by having *E* label as "characteristic" or "not characteristic" different proportions of positive and negative traits. In this case, where a negative concept attitude was being developed, consistency of reinforcement is defined as the ratio formed by the sum of the number of positive traits called "not characteristic" by *E* and the number of negative traits called "characteristic" during concept-attitude development, divided by the total number of traits used for development. Therefore, greatest, or 100%, consistency occurred when all the negative traits were called "characteristic" by *E* and all the positive traits were called "not characteristic." Least, or 50%, consistency occurred when the number of positive and negative traits called "characteristic" by *E* was equal, and the number of positive and negative traits called "not characteristic" was also equal.

Stimulus intensity was varied by using with half of the 80 *Ss* the negative words most often rated "very undesirable" and using with the other half the negative traits most often rated "fairly undesirable." Each of these two stimulus intensity groups was further divided into five sets of *Ss* which received differing degrees of consistency of reinforcement, as follows: 100%, 92%, 80%, 64%, and 50%. This led to 10 subgroups of eight *Ss* each, and each of these 10 subgroups had four women and four men. The statistical model, therefore, is a fixed-constants design with consistency of reinforcement, stimulus intensity, and sex as the bases of classification.

Concept-attitude strength.—After *Ss* had responded to the 50 stimulus traits employed to develop a concept attitude, they were asked to continue responding to 25 further traits which were placed in random order. However, during this series *E* did not say after each prediction whether the trait was characteristic of the people. Responses to these 25 traits were taken as a measure of concept-attitude strength on the assumption that the stronger the negative concept attitude developed, the greater the probability that *S* would continue calling negative traits "characteristic" and positive traits "not characteristic."

Concept-attitude change.—Finally, a third series of 54 randomly ordered traits was presented. The *Ss* were informed that *E* would

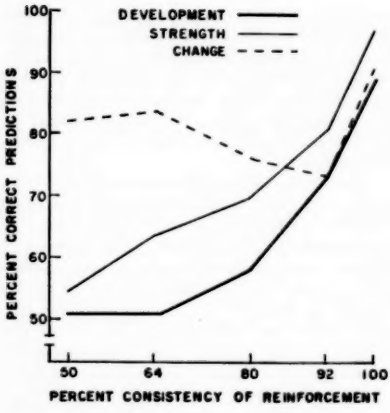

FIG. 1. Concept-attitude development, strength, and change as a function of consistency of reinforcement.

once again tell them after each response whether the trait was characteristic or not characteristic. However, the tactics were reversed during this word series by calling all positive traits "characteristic" and all negative traits "not characteristic." Reactions to these words were used to measure resistance to concept-attitude change.

RESULTS

In effect, *S*s were instructed to predict *E*'s responses during concept-attitude development and change. Therefore, the development and change predictions were scored so that perfect agreement with *E*'s statements would yield the highest score and least agreement would yield the lowest score. Unlike the development and change phases, *E* made no responses when the traits used to measure concept-attitude strength were administered. These traits were employed to determine the degree to which the development experience would lead to predictions consistent with a negative concept attitude. Consequently, the strength measure was scored so that strength scores were highest when *S* called all negative words "characteristic" and all positive words "not characteristic," and lowest when the reverse occurred.

The results for consistency of reinforcement are presented graphically in Fig. 1 and the descriptive statistics are shown in Table 1. An analysis of variance was done for each of the three sets of scores, with Sex, Stimulus Intensity, and Consistency of Reinforcement as the variables of classification. As Table 2 shows, the five Consistency of Reinforcement groups differed beyond the .001 level on the development and strength measures and beyond the .025 level on the measure of change. However, there are no significant main effects for Stimulus Intensity or Sex at the .05 level. According to Lindquist

TABLE 1

MEANS AND *SD*s OF CORRECT RESPONSES FOR THE FIVE LEVELS OF CONSISTENCY OF REINFORCEMENT

Measure	Reinforcement									
	50%		64%		80%		92%		100%	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
Development (50 Traits)	25.50	2.15	25.44	2.02	29.19	3.88	36.63	6.60	44.56	5.52
Strength (25 Traits)	13.50	4.43	16.06	5.71	17.20	4.88	20.25	5.20	24.31	1.93
Change (54 Traits)	44.50	5.52	45.19	5.84	41.19	10.73	39.81	13.12	49.56	2.68

TABLE 2

ANALYSES OF VARIANCE FOR DEVELOPMENT, STRENGTH, AND CHANGE SCORES

Source	*df*	Development *F*	Strength *F*	Change *F*
Consistency of reinforcement (R)	4	54.80**	12.06**	3.32*
Stimulus intensity (I)	1	.85	.12	0.00
Sex (S)	1	.60	.54	3.11
I × S	1	.97	.03	1.03
R × I	4	.76	.73	.84
R × S	4	3.02*	1.65	3.04*
R × I × S	4	.55	.84	.63
Within	60	*19.86*†	*23.20*†	*69.92*†

* $P < .025$.
** $P < .001$.
† Within-cells mean square.

(**12**), the effect on the level of significance due to differences in variances, such as those obtainable from Table 1, is not unduly great. Approximate adjustments can be made by reading the .001 level of significance as .01 and the .025 level as .05.

The development and change measures could also be scored like the strength data. That is, the development data can be scored in terms of *S*'s deviations from a perfect, negative concept attitude, and the change data in terms of deviations from a positive concept attitude. This results in descriptive statistics closely paralleling those in Table 1. Similarly, mean squares of analyses of variance based on this type of scoring are about the same as those leading to Table 2, and the levels of significance remain unchanged.

DISCUSSION

The results of concept-attitude development can be explained by reinforcement theory although alternative interpretations are also possible. The key concepts are reinforcement, competing responses, and the mediating response. It was assumed that because of past social conditioning, positive stimulus words were associated with some mediator, like the mediator representing the verbal response *good*, and negative traits with some other mediator, such as *bad*. Depending upon the consistency of reinforcement, the stimulus provided by each mediator would become associated in different degrees with competing responses, leading to the associations: good → characteristic, good → not characteristic, bad → characteristic, and bad → not characteristic. With relatively more consistency of reinforcement during the development of a negative concept attitude, bad → characteristic and good → not characteristic should be the stronger associations, and bad → not characteristic and good → characteristic should be weaker associations. Consequently, concept-attitude development and strength scores should increase as consistency of reinforcement increases. The data fully support this expectation.

In general, the development and strength data agree with studies of verbal probability learning. It is commonly found (**7, 8, 10**) that the ratio of correct to total responses corresponds approximately with consistency of reinforcement. The development and strength curves indicate a close rank-order relation between the proportion of correct responses and consistency of reinforcement; however, for most instances in the present experiment, the ratio of correct to total responses is smaller than the corresponding reinforcement figure. Goodnow and Postman (**7**) report similar results, and Goodnow (**4**) has discussed several reasons which might account for this phenomenon.

The data for concept-attitude change as a function of consistency of reinforcement show that there is least resistance to change for those *S*s who had the most consistent reinforcement during the initial learning period. This result is similar to what is found in reversal problems used in investigations of acquired distinctiveness of cues (**11**).

Acquired distinctiveness serves in part as an explanation for previous findings on probability learning (**5, 6**). The theory of acquired distinctiveness assumes that a stimulus becomes more distinct as the association between it and a mediating process increases. In the present case the evaluative element of the stimulus traits would become most distinct for the group with the highest level of concept-attitude development. As a result, new learning based upon this distinct stimulus dimension or its discriminated opposite should be facilitated. By this reasoning, the group with the greatest consistency of reinforcement should reverse with relative ease, and that is what was found. A series of *t* tests, using the within variance of the analysis of variance and 60 *df*, indicate that none of the concept-attitude change means of the remaining four groups differed at the .05 level. This suggests that even a little ambiguity is enough to either curtail the development of acquired distinctiveness or keep it below a functionally significant level.

The change results are similar to findings on partial reinforcement of verbal responses. For example, Grant, Hake, and Hornseth (**8**) report a study in which there were groups trained with 100%, 75%, or 50% reinforcement schedules. The 75% and 50% groups were essentially alike and showed greater resistance to change than the 100% group. If their 50% curve is extrapolated to equate the number of predictions made by their *S*s with the number employed during the change stage of the present experiment, there is a close correspondence between the percentage of correct responses in the two studies.

The data from concept-attitude change fail to confirm the seemingly obvious, common sense position which says that an attitude should be more unyielding when a single point of view rather than a mixed view is presented to the learner. Instead, the data appear to confirm the observation that rigidity is associated with ambiguity (**3**) because resistance to change was increased after relatively inconsistent value judgments were reinforced during concept-attitude development.

The results for stimulus intensity were contrary to expectations. Reinforcement theory predicts that reaction potential will be greater with greater stimulus intensity when other factors are equal. However, no statistical difference due to stimulus intensity was found. One explanation is that there may not have been enough difference in intensity between the traits most often rated very undesirable and those most often rated fairly undesirable.

Summary

Attitude was expressed in concept-formation terms, and an empirically derived evaluative dimension was used to study concept-attitude development, strength, and resistance to change. Consistency of reinforcement—the consistency with which negative or positive value-laden characteristics were attributed to a fictional group of people—was found to affect concept-attitude development, strength, and resistance to change. A mediation hypothesis was used to explain these results. Further analysis indicated that concept-attitude development, strength, and resistance to change were neither significantly related to sex nor differentially influenced by two degrees of stimulus intensity.

References

1. Clark, K. B., & Clark, M. Racial identification and preference in Negro children. In G. E. Swanson, T. M. Newcomb, & E. L. Hartley (Eds.), *Readings in social psychology*. New York: Holt, 1952.
2. Eisman, B. S. Attitude formation: The development of a color preference response through mediated generalization. *J. abnorm. soc. Psychol.*, 1955, **50,** 321–326.
3. Frenkel-Brunswik, E. Intolerance of ambiguity as an emotional and perceptual personality variable. *J. Pers.*, 1949, **18,** 108–143.
4. Goodnow, J. J. Determinants of choice-distribution in two-choice situations. *Amer. J. Psychol.*, 1955, **68,** 106–116.
5. Goodnow, J. J., & Pettigrew, T. F. Effect of prior patterns of experience upon strategies and learning sets. *J. exp. Psychol.*, 1955, **49,** 381–389.
6. Goodnow, J. J., & Pettigrew, T. F. Some sources of difficulty in solving simple

problems. *J. exp. Psychol.*, 1956, **51,** 385-392.

7. Goodnow, J. J., & Postman, L. Probability learning in a problem-solving situation. *J. exp. Psychol.*, 1955, **49,** 16-22.
8. Grant, D. A., Hake, H. H., & Hornseth, J. P. Acquisition and extinction of a verbal conditioned response with differing percentages of reinforcement. *J. exp. Psychol.*, 1951, **42,** 1-5.
9. Horowitz, E. L. The development of attitude toward the Negro. *Arch. Psychol. N. Y.*, 1936, No. 194.
10. Humphreys, L. G. Acquisition and extinction of verbal expectations in a situation analogous to conditioning. *J. exp. Psychol.*, 1939, **25,** 294-301.
11. Lawrence, D. H. Acquired distinctiveness of cues: II. Selective association in a constant stimulus situation. *J. exp. Psychol.*, 1950, **40,** 175-188.
12. Lindquist, E. F. *Design and analysis of experiments in psychology and education.* New York: Houghton Mifflin, 1953.
13. Osgood, C. E., & Tannebaum, P. H. The principle of congruity in the prediction of attitude change. *Psychol. Rev.*, 1955, **62,** 42-55.

(Received April 23, 1957)

Journal of Experimental Psychology
Vol. 55, No. 6, 1958

THE RELATION OF VISUAL ACUITY TO CONVERGENCE AND ACCOMMODATION [1]

ENDEL TULVING [2]

Harvard University

Measured visual acuity is usually expressed as the reciprocal of the angle that the minimally visible spatial extent or gap subtends at the observer's eye. This practice implies that the actual size of the acuity target and its distance from *O* are immaterial and that what determines the resolution threshold is the size of the retinal image of the object. Several experiments have, however, been reported showing that there are exceptions to this rule. The evidence appears to show that visual acuity is independent of observation distance when the distances are larger than one or two meters, but that it decreases with decreasing distances below this limit (**1, 5, 7, 8, 9**).

Various explanations have been offered to account for this phenomenon. The most prominent among these has been the suggestion that, since convergence and accommodation necessarily covary with the distance of the test object, either or both of them may be related to this phenomenon (**3, 8, 10**). So far no direct evidence has been available to evaluate this hypothesis. The experiment reported here was designed to provide such information.

[1] This report is based on a part of a dissertation submitted to the Graduate School of Arts and Sciences of Harvard University in partial fulfillment of the requirements for the Ph.D. degree in psychology. The research was supported by the National Science Foundation under grant No. G1280. The author is greatly indebted to E. G. Heinemann and to E. G. Boring for their advice and criticism.

[2] Now at University of Toronto, Canada.

METHOD

Apparatus.—The apparatus consisted of three main parts: a haploscope for varying the convergence of *O*'s eyes, a continuously variable acuity target, and an optical system to provide uniform illumination of the target.

A commercial troposcope, Wottring Model 40 B, manufactured by American Optical Company, was modified to serve as a simple haploscope. The right arm of the haploscope was fixed, so that the right eye looking into it remained always in its primary position. The left arm could be turned both horizontally and vertically about its center of rotation which coincided, more or less, with the center of rotation of *O*'s left eye. The adjustments of the left arm were made to vary the convergence of the left eye. Thus all converging was done by *O*'s left eye only, whereas the right eye in its primary position saw the acuity target.

Each arm of the haploscope was 100 cm long. At the far end of the right arm was a circular opening through which the acuity target could be seen when exposed. A small bright dot, about 1 mm. in diameter, was reflected from a cover glass inside the arm and served as the right fixation point. A similar fixation point was also placed at the far end of the left arm. When the target was not exposed but both fixation points were turned on, the *O* looking into the instrument saw a totally dark field with two fixation points. When the convergence of *O*'s left eye corresponded to the setting of the left arm of the haploscope, the two fixation points were seen fused as one.

The acuity target consisted of two vertical parallel black lines seen against an evenly illuminated circular field 17.5 mm. in diameter which, at a distance of 100 cm., corresponds to visual angle of 1°. The width of the black lines, formed by two very taut wires, was .014 in. or 1 min. 15 sec. in angular measure. The distance between the two lines could be continuously varied by *E* and read to the nearest 10^{-4} in., corresponding to approximately $\frac{1}{2}$ sec. of arc.

The optical system consisted of a ribbon filament lamp, a pair of condensing lenses, a filter holder for the insertion of neutral density filters, a circular shutter disk, and a flashed opal screen. The acuity target was seen against

this screen. The luminance of the screen, calibrated by means of a MacBeth Illuminometer, was kept constant at 19 mL. The exposure time of the target was controlled by the circular shutter and was constant at .25 sec. throughout the experiment. Various stops and screens were used in the optical system, as well as in the arms of the haploscope, to keep out stray light.

The eyes of a normal young *O*, owing to convergent accommodation, are accommodated for the convergence distance (**4**). To keep the target in good focus for all amounts of convergent accommodation, very small artificial pupils, .5 mm. in diameter, were employed.

Observers.—Seven men, aged from 22 to 29 yr., with a median age of 23 yr., served as *O*s. Six of them had no known visual defects, the seventh one was a slight myope.

Task.—The *O* sat in a light-tight but ventilated cubicle. His head was supported by the headrest of the haploscope. Looking into the instrument, his task was to fuse the two separate fixation points and then present the acuity target to himself by pressing a key. This key operated the shutter permitting the circular background of the two black wires to be illuminated for .25 sec. The *O* reported whether he saw one or two black lines by operating two other keys within his easy reach. The intervals between successive exposures were about 8 to 10 sec. Prior to the experiment proper all *O*s received sufficient practice in fusing the two fixation points under all degrees of convergence used in the experiment.

Procedure.—In Part 1 of the experiment each *O* came for three experimental sessions. These constituted repetitions of the same procedure. In each session, lasting about 2 hr., including frequent rest periods, *O*'s acuity was measured for five different angles of convergence which corresponded to convergence distances of 25 cm., 33 cm., 50 cm., 100 cm. and infinity.

To determine thresholds of separation between the two black lines of the acuity target the so-called "up-and-down" method was used (**2**). The *O*'s threshold for each convergence angle in a given session was based upon 60 judgments. The order of convergence angles within sessions was counterbalanced.

Two *O*s made their observations in the first three sessions while their ciliary muscles were paralyzed by homatropin. Prior to each session, drops of 5% homatropin solution were applied to the conjunctival sacs of both eyes at 15-min. intervals. Applications were stopped when retinoscopic measurements revealed the refraction of the right eye to be the same for the convergence at 25 cm. as at 6 m.

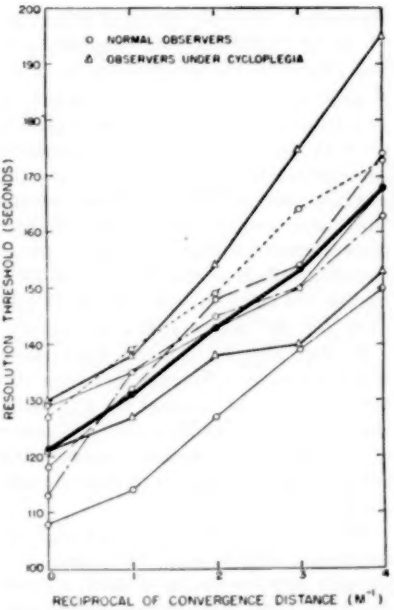

FIG. 1. Threshold separations of the minimally separable two black lines as a function of convergence. Means of individual unpracticed *O*s. The heavy line shows the grand mean.

In Part 2, following the initial three sessions, four of the seven *O*s were given daily practice periods of about 1 hr. in viewing and making judgments about the acuity target under the two extreme angles of convergence used previously. Two *O*s received such practice for 12 days, the other two for 15 days.

After the conclusion of the practice sessions the acuity of the four *O*s was again measured for the original five angles of convergence under the same experimental conditions as in the first part of the experiment.

RESULTS

The results of Part 1, in which seven relatively inexperienced *O*s were employed, are shown in Fig. 1. Threshold separation of the two parallel lines is plotted against the reciprocal of the convergence distance for all *O*s. Each point on the graph

represents the over-all mean threshold of an *O* based on 180 judgments of the target in three different sessions. A systematic change in the threshold as a function of convergence is clearly present in all *O*s. The threshold increases with the decrease in the convergence distance. This means that the acuity is less at shorter convergence distances. The mean threshold of all seven *O*s is about 40% higher for the convergence at 25 cm. than for the convergence at infinity.

It will also be noticed that there is no essential difference between the results from the normal *O*s and from the two *O*s under the influence of homatropin. Acuity varies with convergence whether accommodation changes or not.

The results of Part 2, in which four relatively experienced *O*s were employed, are shown in Fig. 2. Each point on the graph represents the mean threshold of an *O* based on 180 judgments of the target in three sessions following 12 to 15 hr. of practice. By and large the effect found with unpracticed *O*s is now gone: the threshold of separation is essentially the same for all degrees of convergence. For two *O*s there still remains a small effect similar to that found with unpracticed *O*s—the threshold is somewhat higher for smaller convergence distances. But even for these two *O*s the magnitude of the effect is no greater than about 4%.

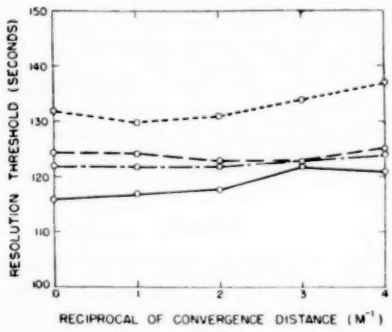

FIG. 2. Threshold separations of the minimally separable two black lines as a function of convergence. Means of individual practiced *O*s.

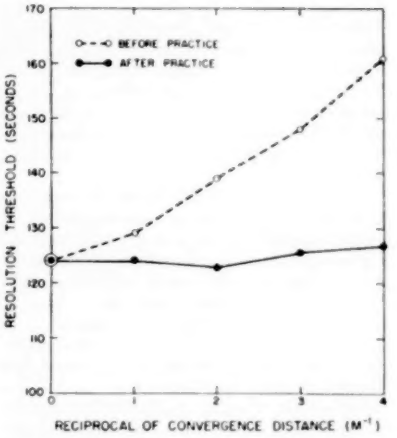

FIG. 3. Practiced vs. unpracticed *O*s ($N = 4$). Threshold separations of the minimally separable two black lines as a function of convergence.

In Fig. 3 are shown the mean thresholds of the four *O*s in both Part 1 and Part 2, following the practice period. The difference between the two curves is obvious.

DISCUSSION

Freeman's "anomaly of visual acuity" (**5**) has remained a puzzle for over a quarter of a century: visual acuity is less for very close targets than for far. Although there is no known mechanism by which accommodation or convergence might influence visual acuity these two oculomotor adjustments have been brought in to explain the phenomenon (**3, 8, 10**). So far, to the best knowledge of the writer, no experiments have been

reported in which convergence and accommodation have been isolated experimentally and their effect on acuity observed. In acuity-distance experiments the distance of the test object, convergence, and accommodation all vary together and this leaves the door open to the argument that "psychic" factors are responsible for the phenomenon (**11**, p. 495).

The findings of the present experiment indicate that convergence alone, in absence of any changes in the distance of the acuity target and regardless of the state of accommodation, influences acuity. Thus it seems that changes in convergence may have produced the results similar to Freeman's "anomaly" in previous acuity-distance experiments. The previous results, however, can be regarded as anomalous only as far as the basic principles of physiological optics are concerned. The present experiment suggests that the anomaly disappears when the past history of the *O* making acuity judgments is taken into consideration. The *O*s with some skill in the task yield results which are in good agreement with the principles of physiological optics.

It is, of course, quite possible that the findings of this experiment cannot be generalized to account for discrepancies between theory and empirical facts in the acuity-distance experiments. But the hypothesis that experienced and inexperienced *O*s would give different results in an experiment such as Giese's (**7**) is directly testable and may be well worth considering.

The present investigation does not permit any explanation as to why practice in the experimental task should affect acuity under greater degrees of convergence and not under convergence at infinity. It is a well established fact that most psychophysical thresholds, including visual acuity, can be changed by practice (**6**), but the mechanisms underlying it are not too well understood. A precise analysis of what has been here loosely called "practice" might be helpful in delineating the direction in which the answer should be sought.

At present it is only reasonable to assume that practice does not affect the power of retinal resolution. Thus the differences in acuity for close targets between practiced and unpracticed *O*s are probably related to the observational procedures that these *O*s use. It looks as if something—we cannot yet say what it is—confuses the unpracticed *O* so that he cannot for near objects, or with greater degrees of convergence, take the same advantage of his power of retinal resolution as he can for far objects, or with small angles of convergence.

Summary

In this experiment the acuity of the *O*s' right eye in its primary position was measured while the left eye changed its convergence. The acuity target, at a constant distance from *O*, consisted of two parallel lines.

Acuity was found to decrease with increasing angles of convergence when relatively unpracticed *O*s were used. This effect was shown to be independent of changes in convergent accommodation. For *O*s who had received longer practice in the experimental situation, acuity was found to be independent of both convergence and convergent accommodation. The results were discussed with reference to the previous experimental findings that measured visual acuity decreases with observation distance for relatively small distances.

References

1. Beebe-Center, J. G., Mead, L. C., Wagoner, K. S., & Hoffman, A. C. Visual acuity and distance of observation. *J. exp. Psychol.*, 1945, **35**, 473–484.
2. Dixon, W. J., & Mood, A. M. A method for obtaining and analyzing sensitivity data. *J. Amer. statis. Ass.*, 1948, **43**, 109–126.
3. Falk, J. L. Theories of visual acuity and their physiological bases. *Psychol. Bull.*, 1956, **53**, 109–133.
4. Fincham, E. F. The proportion of ciliary muscular force required for accommodation. *J. Physiol.*, 1955, **128**, 99–112.
5. Freeman, E. Anomalies of visual acuity in relation to stimulus distance. *J. opt. Soc. Amer.*, 1932, **22**, 285–292.

6. Gibson, E. J. Improvement in perceptual judgments as a function of controlled practice and training. *Psychol. Bull.*, 1953, **50**, 401–431.

7. Giese, W. J. The interrelationship of visual acuity at different distances. *J. appl. Psychol.*, 1946, **30**, 91–106.

8. Luckiesh, M., & Moss, F. K. The variation in visual acuity with fixation-distance. *J. opt. Soc. Amer.*, 1941, **31**, 594–595.

9. Musylev, F. J. Die Abhängigkeit der Sehschärfe von der Entfernung des Objektes. *Acta Ophthal., Kob.*, 1937, **15**, 216–226.

10. Sloan, L. L. Measurement of visual acuity. *A. M. A. Arch. Ophthal.*, 1951, **45**, 704–725.

11. Walls, G. L. Factors in human visual resolution. *J. opt. Soc. Amer.*, 1943, **33**, 487–505.

(Received June 3, 1957)

Journal of Experimental Psychology
Vol. 55, No. 6, 1958

REPRODUCTIVE INTERFERENCE FOLLOWING "APPROPRIATE" AND "INAPPROPRIATE" WARM-UP ACTIVITIES

LELAND E. THUNE

Vanderbilt University

This study had its origin within two different areas of verbal-learning behavior. One is the role of set and context upon reproductive interference (RI). The other is the role of the warm-up effect in verbal learning and retention.

An experiment by Jenkins and Postman (**8**) called attention to the role of set as a factor influencing amount of RI. Different response sets were established by using different testing procedures during the learning of the original and the interpolated tasks. Retention for the original material was subsequently tested for different groups of *S*s by each testing procedure. "When both activities were carried out under the same set, a greater amount of retroactive interference resulted than was found when original learning and interpolated learning took place under different sets. Like dissimilarity of materials, dissimilarity of sets contributes to the functional isolation of tasks learned in succession" (**8**, p. 72). Similar results have been obtained by Postman and Postman (**9**) and others.

Likewise, several studies within the framework of the RI paradigm have shown that RI is reduced when the original learning and the interpolated learning are carried out in different environmental contexts. For example, Bilodeau and Schlosberg (**1**) found that when the interpolated material was learned in a different room and with a different type of exposure device than was employed during the learning of the original material, scarcely half as much RI was obtained (in comparison to a control group) as when both the original and the interpolated learning were carried out in the same room and with the same exposure device. This finding was confirmed and extended in a study reported recently by Greenspoon and Ranyard (**2**) which investigated the effect of environmental and procedural context upon RI under various combinations of "similar" or "different" contexts for the original learning, interpolated learning, and relearning.

The role of warm-up effects in the retention of verbal materials has not been explored intensively, although several studies (**6, 7**) have shown that participating in a set maintaining or reinstating warm-up activity may facilitate the retention and relearning of previously learned materials. However, it should be noted that studies by Rockway and Duncan (**10**) and by Hovland and Kurtz (**4**) have failed to show evidence of such an effect under conditions which might have been expected to produce it.

There is a growing bibliography of studies concerned with warm-up effects in verbal learning, but little is known about the fundamental nature of the warm-up phenomenon. Warm-up effects are usually discussed in terms of set, but it has been equally difficult to specify the fundamental nature of set, and the relative dimensions along which it may be varied.

Warm-up effects have sometimes

been thought of in terms of peripheral factors, such as postural adjustments, or re-adaptation to the rhythm of the exposure device (**5, 11**). For example, Irion suggests that ". . . certain postures and muscular tension patterns may be optimal for the performance of a given task. To the extent that these optimal postures and tensions are acquired during original learning and are disrupted between the time of original learning and recall, there may be a decrement in performance which is attributable to this decreased state of responding efficiency" (**5,** p. 340).

At other times, warm-up in verbal learning is treated in terms of more central factors, such as attentive adjustments and ". . . the establishment of a 'readiness' set to make a response just as soon as the next stimulus appears" (**12,** p. 255). Underwood has obtained evidence that color naming, as a rest interval activity, may induce ". . . a set to respond to stimuli more frequently than would be the case without color naming . . ." (**13,** p. 339). Although Underwood's study was not directed toward the study of warm-up effects, his finding appears relevant here since color naming has been employed frequently as a warm-up activity.

The aim of the present study was to determine whether, in an RI situation involving different learning contexts for the original and interpolated tasks, a warm-up activity specifically appropriate to the *interpolated* learning context will have an effect upon the retention and relearning of the original material different from that of a warm-up activity appropriate to the context and conditions present during the learning (and later relearning) of the *original* material. The results should bear upon the question as to the relative roles of central and peripheral factors in verbal warm-up. If warm-up is essentially peripheral in nature, participation in a warm-up task administered within the interpolated learning context should fail to facilitate, if not interfere outright with, the recall of the original material. On the other hand, if warm-up effects in verbal learning are a predominantly central phenomenon, warm-up practice within either the original or the interpolated learning context should facilitate the recall and relearning of the original material, since essentially the same central factors should be common to both warm-up situations.

Method

Subjects.—Seventy-eight undergraduate students served as *Ss* in fulfillment of a course requirement in general psychology. The *Ss* were randomly assigned to one of three conditions, 26 to each condition.

Apparatus.—Two types of paired-associate exposure devices were employed. One was a pair of traditional Hull-type memory drums operating on a 4-sec. cycle, the stimulus item alone being exposed for the first 2 sec. of the cycle, the stimulus and response items being jointly exposed during the last 2 sec. Successive trials were separated by an 8-sec. interval. The *S* responded verbally. Two separate memory drums were employed in order to eliminate the necessity of changing tapes either within, or between, sessions. The two memory drums were mounted side by side within a plywood framework which shielded *E* from *S*'s view.

The second paired-associate exposure device was a filmstrip projector equipped with an automatic film advancing mechanism. The stimulus and response material was prepared on filmstrips and was projected onto a small white screen about 8 ft. from the projector. A 5-sec. exposure cycle was employed; the stimulus item alone being exposed for 3 sec., and the stimulus and response items being jointly exposed during the remaining 2 sec. Successive trials were separated by a 10-sec. rest interval. The *S* sat at a table about 4 ft. in front of the projection screen and made his responses by writing them on small sheets of ruled paper held on a clipboard, a separate sheet being employed on each trial. The *S* held a piece of

cardboard in his nonwriting hand, and after writing a response word, or making a dash line to indicate his inability to make a response, *S* adjusted the cardboard so as to conceal all previous responses before attempting the next anticipation. The *E* stayed near the filmstrip projector, which was located about 4 ft. behind *S*. From this position *E* could observe *S* inconspicuously to insure that he used the masking card correctly, and that he did not write responses after they had been projected onto the screen.

Experimental design.—The experimental design was a modification of the traditional RI paradigm, and is shown schematically in Table 1. This same table also shows the performance criterion employed in each phase of the experiment. The experimental variable which served to distinguish the several conditions was the type of warm-up activity which *S* engaged in immediately prior to relearning the original list (List A) on Day 3.

Each *S* served in three individual sessions, the three sessions for any given *S* always falling on consecutive days, but not necessarily at the same hour on each day.

List A was presented on a Hull-type memory drum during both its original learning on Day 2 and its relearning on Day 3. The interpolated list (List B) was always presented by means of the filmstrip projector. Thus the method of presenting the stimulus material, *S*'s method of making a response, the duration and rhythm of the exposure cycle, and *S*'s posture and his position within the experimental room all differed during the interpolated learning from what they had been during the original learning.

Prior to relearning List A on Day 3 all *Ss* participated in a warm-up, or set-reinstating, activity within either the memory drum context in which List A was originally learned (and was to be subsequently relearned) or the filmstrip context in which List B had been learned.

The *Ss* assigned to Cond. II received 10 trials on a warm-up list administered on the memory drum used with List A. Hence their warm-up practice can be thought of as being appropriate to the recall of that list. The *Ss* assigned to Cond. III likewise received 10 trials of warm-up practice immediately prior to relearning List A, but for this group the warm-up task was carried out within the filmstrip learning context in which List B was learned, and hence can be thought of as being appropriate to the recall of the interpolated list. The *Ss* serving under Cond. I, the control condition, also received 10 trials on one or the other of these activities. However, for this group, the task was performed at the end of Day 2, and merely served to insure that any long-term practice or motivational effects resulting from participation in the warm-up task would operate equally among all three groups. Half of the *Ss* in the control group were given the memory drum warm-up task used with Cond. II, and the other half received the filmstrip warm-up task employed with Cond. III.

Verbal lists.—Two practice lists were employed; one was presented on a memory drum and the other was presented by means of the filmstrip projector. Each list was composed of eight pairs of unrelated one-syllable nouns, different words being used on the two lists.

List A was composed of 12 pairs of unrelated

TABLE 1

DESIGN OF THE EXPERIMENT

Condition	Day 1		Day 2			Day 3	
I. Control $N = 26$	Practice List Memory Drum	Practice List Filmstrip	List A Memory Drum	List B Filmstrip	Either {Warm-up List Memory Drum / Or {Warm-up List Filmstrip		Relearn List A Memory Drum
II. OL Warm-up $N = 26$	Ditto	Ditto	Ditto	Ditto		Warm-up List Memory Drum	Ditto
III. IL Warm-up $N = 26$	Ditto	Ditto	Ditto	Ditto		Warm-up List Filmstrip	Ditto
Criterion to which practice was carried	Two successive errorless trials or minimum of 10 trials		9 correct in 12-item list	9 correct in 12-item list	10 trials	10 trials	Two successive errorless trials or minimum of 10 trials

TABLE 2

MEAN TRIALS TO REACH TWO SUCCESSIVE ERRORLESS TRIALS ON THE PRACTICE LISTS, AND 9/12 CORRECT ON LISTS A AND B

Condition	Practice List Memory Drum		Practice List Filmstrip		List A (OL) Memory Drum		List B (IL) Filmstrip	
	Mean	σ_M	Mean	σ_M	Mean	σ_M	Mean	σ_M
I	18.00	1.79	7.65	1.10	9.46	.75	6.85	.74
II	18.81	1.84	6.54	.67	10.08	.76	6.54	.69
III	17.69	1.44	6.38	.49	9.73	.88	6.69	.49

two-syllable adjectives. It was typed in capital letters on glazed white cloth tape for presentation on a memory drum.

List B was likewise composed of 12 pairs of two-syllable adjectives, and was related to List A in an A-B, A-K manner. List B was prepared for presentation through the filmstrip projector.

The usual rules governing the construction of paired-associate lists of low association value were observed in preparing the lists, and all lists were presented in five different serial orders.

Warm-up task.—The warm-up task presented *S* with an activity which had many procedural and contextual similarities either to the original or to the interpolated learning situations, yet which provided minimal opportunity for rehearsal or learning-to-learn. It was rather similar to that employed in a previous study (11). The warm-up list was prepared in two forms, one for use on a memory drum, the other for use with the filmstrip projector. In other respects the two lists were identical. Although the lists were prepared in paired-associate form, the identical stimulus figure (a bold horizontal line) was used for all pairs. The response item was a geometrical figure, either a cross, a square, a triangle, a diamond, or a circle. The *S* was required to guess, each time the stimulus figure came into view, which one of the five geometrical response figures would appear next, and to respond with his guess *before* the response figure came into view.

The warm-up lists were presented in exactly the same manner as the corresponding verbal lists, i.e., the warm-up list employed with Cond. II was presented on the same memory drum and in the same manner as was List A. Similarly, the warm-up list used with Cond. III was prepared on a filmstrip, and was presented using the same procedure as was employed with List B. Both forms of the warm-up list were prepared in five different serial orders, the sequence and frequency with which each of the five response figures appeared in each serial order being determined by use of a table of random numbers.

The instructions for the warm-up task in no way linked it with the previous learning task, it being referred to simply as a "study in guessing." Furthermore, at the time *S* participated in the warm-up activity, he did not know that he was going to be required to relearn List A.

RESULTS

Initial comparability of groups.—The three groups of *S*s received identical treatment up through the learning of List B on Day 2. As shown in Table 2, the groups were highly comparable in the rates with which they learned the two practice lists, List A and List B. Analysis of variance tests based upon the number of trials required to reach the specified criterion on each of these four lists confirmed the comparability of the groups. None of the resulting *F* ratios approached significance.

Relearning on List A.—Three performance measures were employed in evaluating the effect which the two different types of warm-up activities

TABLE 3

RELEARNING LIST A

Cond.	Number Correct on RL Trial 1		Number Correct on RL Trials 1–10		Number of Trials to Criterion	
	Mean	σ_M	Mean	σ_M	Mean	σ_M
I	3.81	.23	89.69	2.62	11.73	1.09
II	5.62	.33	100.46	1.70	6.92	.45
III	4.00	.38	96.88	2.11	7.46	.55

had upon the retention and relearning of List A. Table 3 summarizes the relearning data. The relearning curves for List A on Trials 1 through 10 are shown in Fig. 1.

Performance on the first relearning trial on List A was analyzed by means of an analysis of covariance in order to take account of any small initial differences between groups, even though these differences had been previously shown to lack statistical significance. The total number of correct anticipations on the first four original learning trials on List A was used as the initial measure in the analysis of covariance since some *Ss* received only four trials, having attained the original learning criterion of 9 correct on the fourth trial. The analysis of covariance yielded an F of 12.87 which is significant at the .001 level.

Performance on the first 10 relearning trials taken together was likewise analyzed by means of an analysis of covariance in which the number of correct anticipations on the first four original learning trials was employed as the initial measure. The resulting F of 9.56 is significant at the .001 level.

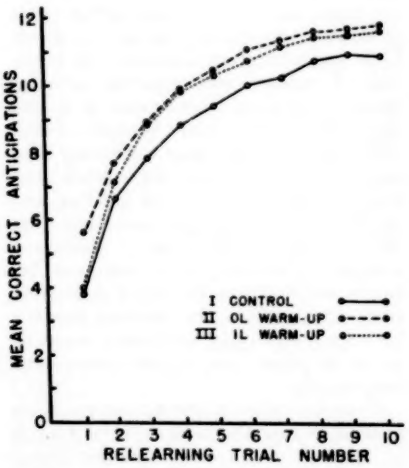

FIG. 1. Mean number of correct anticipations per trial on Trials 1-10 in relearning List A.

TABLE 4

FREQUENCY OF INTERLIST INTRUSIONS FROM LIST B DURING RELEARNING OF LIST A

Condition	Number of *Ss* Who Gave One or More Intrusions	Total Number of Intrusions by All *Ss*
I. Control	9	16
II. OL Warm-up	6	14
III. IL Warm-up	12	29

The raw trials to reach two successive errorless repetitions were subjected to a logarithmic transformation in order to eliminate a significant departure from homogeneity of variance. The transformed scores were then analyzed by an analysis of covariance in which the number of trials to reach 9 out of 12 correct during the original learning of List A was employed as the initial measure. The resulting F of 11.39 was significant at the .001 level.

Data regarding the frequency of intrusions from List B during the relearning of List A are summarized in Table 4. Although only half as many *Ss* in Cond. II gave one or more intrusions as did *Ss* under Cond. III, the number of *Ss* in each of the three conditions who gave one or more intrusions did not differ significantly when tested by a chi-square test ($\chi^2 = 3.059$, $\chi^2_{.10} = 4.605$ for 2 *df*).

DISCUSSION

The results demonstrate clearly that a warm-up effect may be present during the relearning of verbal materials. All three measures employed in evaluating relearning performance on List A confirmed the significance of the difference between groups, differences which, it

would appear, can be accounted for only through the postulation of some such factor as has been called a warm-up effect. In this respect, the results support the findings of Irion (**6**) and of Irion and Wham (**7**) on warm-up effects in retention, and parallel those of Thune (**11, 12**) and Hamilton (**3**) on warm-up effects in original learning.

We are unable to account for the negative findings of Rockway and Duncan (**10**) and of Hovland and Kurtz (**4**) with respect to warm-up effects in retention. However, there was at least one important difference between the warm-up tasks employed by those investigators, as well as by Irion (**6**) and Irion and Wham (**7**), and the one employed in the present study. Previous studies on warm-up effects in verbal retention have employed color *naming* (i.e., reading off "response" colors after they come into view), whereas the present study, as well as a previous study by Thune (**11**) on warm-up in original learning, required *anticipatory guessing* (i.e., making an anticipatory guess as to which "response" color or figure would appear next). Perhaps the anticipatory guessing type of warm-up task is more effective than the naming task in developing the kind of anticipatory response set demanded for efficient recall and relearning.

From Fig. 1 it can be seen that the three groups maintained the same rank order throughout the first 10 relearning trials, Cond. I consistently ranking the lowest and Cond. II consistently ranking the highest. However, whereas the relearning curves for Cond. I and II parallel each other with rather similar slope values, the curve for Cond. III starts out on the first relearning trial at approximately the same level as Cond. I but rises with a greater initial slope than either of the other groups so that by the third trial this group is performing at a level close to that of Cond. II, a position which it maintains throughout the remainder of the first 10 relearning trials. The difference between Cond. I and III is not significant on the first relearning trial ($t = .42$), whereas the difference between each of these groups and Cond. II is significant (Cond. I and II, $t = 4.44$; Cond. III and II, $t = 3.19$; $t_{.01} = 2.68$). However, on the third relearning trial Cond. II and III do not differ significantly ($t = .25$), but the difference between each of these groups and Cond. I was significant at a moderate level of confidence (Cond. I and II, $t = 2.21$; Cond. I and III, $t = 1.96$; $t_{.05} = 2.01$).

The failure of Cond. III to show significant warm-up facilitation on the first relearning trial, although showing it clearly on later trials, warrants some comment or attempt at explanation. It is suggested that relearning performance under Cond. III was influenced by both a facilitating and an inhibiting warm-up effect. The apparent lack of warm-up on the first relearning trial may merely reflect the net effect of two opposing warm-up processes, one facilitating and the other inhibiting the recall and relearning of List A.

In experiments involving both positive and negative transfer it has been commonly observed that negative transfer effects show themselves most prominently during the early stages of learning (or relearning), and that as practice continues the net transfer effect tends to become less negative or more positive. It is suggested that just such an interplay of positive and negative warm-up factors may have been operating during the relearning of List A under Cond. III. Thus, on the first relearning trial there may have been appreciable, but roughly equal, amounts of positive and negative warm-up present, the net effect being essentially zero. The rapid increase in relearning performance between the first and the third relearning trials may reflect the shifting balance between a relatively transitory negative warm-up effect, and a more permanent positive effect.

If we conceive of the warm-up effects present during the relearning of List A as having both *central* and *peripheral* components, and if we grant that these two processes may have different strengths relative to one another at

different stages in the relearning process, we can develop an explanation which, while admittedly *post hoc*, seems to fit the present situation quite adequately.

It is suggested that the peripheral element within the warm-up process functions only during the first few relearning trials, whereas the effect of the central component endures at least throughout the first 10 relearning trials. Under Cond. II, both types of warm-up would serve to facilitate the relearning of List A, since the warm-up task was administered within the same environmental context in which List A was originally learned and subsequently relearned. However, under Cond. III the warm-up task was introduced within the interpolated learning context and the peripheral element of the two warm-up factors might be expected not only to fail to prepare *S* for efficient recall and relearning of List A, but even to produce active competition from List B. However, the central component of the warm-up effect developed under Cond. III might be expected to facilitate List A relearning, since it should be compatible with and supportive of efficient recall and relearning.

In summary, then, it is suggested that warm-up effects in the retention of verbal materials have both a central and a peripheral component, and that these components may have varying strengths relative to one another at different phases in relearning. Whether each of these components facilitates or inhibits subsequent recall and relearning depends upon the compatibility between the central and peripheral processes induced by the warm-up activity and the central and peripheral demands of the recall and relearning task.

It would appear that warm-up effects in the retention of verbal materials are more central than peripheral in nature under the conditions of the present study. As we have noted, the warm-up task for Cond. III was carried out in a very different context than was the original learning (and relearning) of List A. The *S*'s posture and position within the room, his mode of making a response, and the time cycle and rhythm of the exposure device with its accompanying noises, were all very different during the warm-up activity from what they were during the original learning and subsequent relearning of List A. Yet it is evident from Fig. 1 that, apart from the first two relearning trials, the two different types of warm-up activities represented by Cond. II and III facilitated the relearning of List A to almost equal degrees, and both activities resulted in significant facilitation of relearning when compared with Cond. I.

Summary

Three groups of 26 *S*s partially learned two related paired-associate lists, which were presented within widely differing procedural and environmental contexts. The groups differed in the amount and type of figure-guessing warm-up practice which they received immediately prior to relearning the original list on the next day. The *S*s in Cond. I (Control) received no effective warm-up practice; those in Cond. II (OL Warm-up) received warm-up practice in the context within which the original list was learned; and those in Cond. III (IL Warm-up) received warm-up practice within the interpolated learning context.

Warm-up practice within the context in which the original list was learned markedly facilitated its relearning. However, warm-up practice administered within the context in which the interpolated list was learned failed to produce a significant effect during the first two relearning trials on the original list, but showed a consistent facilitating effect on all trials thereafter.

It is suggested that the warm-up effect in verbal learning has both peripheral and central components, but that the peripheral component is more transitory in its effect.

References

1. Bilodeau, I. McD., & Schlosberg, H. Similarity in stimulating conditions as a variable in retroactive inhibition. *J. exp. Psychol.*, 1951, **41**, 199–204.
2. Greenspoon, J., & Ranyard, R. Stimulus conditions and retroactive inhibition. *J. exp. Psychol.*, 1957, **53**, 55–59.
3. Hamilton, C. E. The relationship between length of interval separating two learning tasks and performance on the second task. *J. exp. Psychol.*, 1950, **40**, 613–621.

4. Hovland, C. I., & Kurtz, K. H. Experimental studies in rote-learning theory: IX. Influence of work-decrement factors on verbal learning. *J. exp. Psychol.*, 1951, **42,** 265–272.
5. Irion, A. L. The relation of "set" to retention. *Psychol. Rev.*, 1948, **55,** 336–341.
6. Irion, A. L. Retention and warming-up effects in paired-associate learning. *J. exp. Psychol.*, 1949, **39,** 669–675.
7. Irion, A. L., & Wham, D. S. Recovery from retention loss as a function of amount of pre-recall warming-up. *J. exp. Psychol.*, 1951, **41,** 242–246.
8. Jenkins, W. O., & Postman, L. An experimental analysis of set in rote learning: Retroactive inhibition as a function of changing set. *J. exp. Psychol.*, 1949, **39,** 69–72.
9. Postman, L., & Postman, D. Change in set as a determiner of retroactive inhibition. *Amer. J. Psychol.*, 1948, **61,** 236–242.
10. Rockway, M. R., & Duncan, C. P. Pre-recall warming-up in verbal retention. *J. exp. Psychol.*, 1952, **43,** 305–312.
11. Thune, L. E. The effect of different types of preliminary activities on subsequent learning of paired-associate material. *J. exp. Psychol.*, 1950, **40,** 423–438.
12. Thune, L. E. Warm-up effects as a function of level of practice in verbal learning. *J. exp. Psychol.*, 1951, **42,** 250–256.
13. Underwood, B. J. Studies of distributed practice: IV. The influence of rest-interval activity in serial learning. *J. exp. Psychol.*, 1952, **43,** 329–340.

(Received June 3, 1957)

Journal of Experimental Psychology
Vol. 55, No. 6, 1958

THE RELATION OF EYE MOVEMENTS, BODY MOTILITY, AND EXTERNAL STIMULI TO DREAM CONTENT[1]

WILLIAM DEMENT[2] AND EDWARD A. WOLPERT

Department of Physiology, University of Chicago

The phenomenon of dreaming has always occupied an important place in the study of human behavior. Questions such as, When do dreams occur? How long do they last? Does everyone dream? and many others have been the subject of much lively debate. Only recently, however, has it become possible to formulate answers to some of the more fundamental problems of dreaming in objective terms.

The first step in answering these questions was made by Aserinsky and Kleitman (1) who observed periods of rapid, conjugate eye movements during sleep and tested the possibility that such eye movements were related to dreaming. These investigators found that *Ss* awakened when rapid eye movements were present showed a high incidence of dream recall, and that awakenings when the eyes were quiescent elicited a very low incidence of dream recall. These findings were subsequently confirmed in both normal and schizophrenic *Ss* (6), and a recent study (7) appears to have conclusively established rapid eye movements as an objective criterion of dream activity. In the latter study, an 80% incidence of vivid, detailed dream recall was elicited from 9 *Ss* after nearly 200 awakenings during periods of rapid eye movements over many nights of sleep. Only 7% recall was obtained after a comparable number of awakenings, randomly interspersed, when the eyes were quiescent. Of the 7%, the majority of instances of recall occurred when the awakenings followed the cessation of rapid eye-movement periods by less than 8 min., suggesting that such recall represented persisting memory of the dream related to the preceding eye-movement period, rather than dreaming during the period of ocular quiescence. In addition, it was demonstrated that the subjective estimate of dream duration and the amount of dream narrative correlated highly with the length of the rapid eye-movement period before awakening, and that the eye movements themselves seemed to be meaningfully related to the visual imagery of the dream.

By continuous recording of the EEG and eye-movement potentials of 33 adults during 71 nights of undisturbed sleep (8), it was found that eye movements appeared without exception on each of the 71 nights and occurred regularly in relation to the lightest phases of a cyclic variation in depth of sleep as measured by brain-wave patterns. Figure 1, a composite diagram of a single night of sleep constructed from these data, shows both the typical depth of sleep curve as measured by the EEG and also the average durations and times of occurrence of successive eye-movement (dream) periods.

[1] Aided by a grant from the Wallace C. and Clara A. Abbott Memorial Fund of the University of Chicago. The authors wish to acknowledge the helpful criticism of Nathaniel Kleitman during the preparation of this paper.

[2] Postdoctoral Public Health Service Research Fellow of the National Institute of Neurological Diseases and Blindness. Present address, Mount Sinai Hospital, New York.

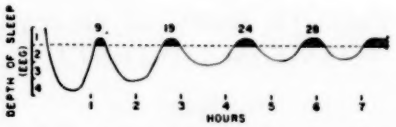

FIG. 1. A typical night of sleep. The depth of sleep in terms of brain waves fluctuates regularly. Whenever it reaches the Stage 1 level, dreaming (shaded areas) starts, accompanied by rapid eye movements, and continues until the downswing of the cycle. The numbers over the shaded areas indicate the average length of successive eye-movement (dream) periods.

In the studies to be described, several aspects of dreaming have been investigated with the aid of eye-movement and brain-wave recordings. In addition to obtaining information pertaining to specific problems, it was hoped that some of the advantages and shortcomings of this technique would be clarified.

METHOD

The *Ss* were 14 men and 2 women. The same general procedure was followed throughout. The *S* reported to the laboratory a little before his accustomed bedtime. Two or three electrodes were affixed to appropriate areas of the scalp with collodion for registering brain waves and two or more electrodes were attached near the eyes to pick up eye-movement potentials. This electrical method of recording has been used by others to investigate the effect of drugs (**2**), stress (**13**), and diurnal variations (**11**) on ocular coordination. In these studies, it was assumed that when the eyes moved, the spatial changes of their associated corneo-retinal potential fields with reference to the stationary electrodes, generated the observed potentials. The criteria of eye-movement potentials during sleep and their differentiation from brain-wave artifacts have been discussed at length elsewhere (**1, 6**).

The *Ss* slept in a quiet dark room. The lead wires, having been gathered together at the top of the head, ran as a single cord to the lead box. This arrangement allowed a full range of movement without entanglement and minimized any annoyance to *S*.

A Grass Model III Electroencephalograph in an adjacent room was used to amplify and record the EEG patterns and eye-movement potentials. The machine was run continuously during the night except when *Ss* were aroused. At these times, recording was interrupted in order to avoid excessive movement and muscle artifacts.

As a rule, *Ss* were awakened during each successive eye-movement period (see Fig. 1). In this way, four or five dream narratives were obtained in a single night. Because it was impossible to predict how long eye-movement periods might last, they were usually interrupted by an awakening from 5 to 20 min. after their onset, values somewhat less than their average duration in undisturbed sleep. The ringing of a doorbell, sufficiently loud to ensure an abrupt termination of sleep, was used as the arousing stimulus throughout these experiments. Upon awakening, *Ss* described the content of their dreams, speaking into a tape recorder or dictaphone placed near the bed. The recordings were later transcribed for further analysis and study. A total of 204 transcriptions was accumulated including some obtained in a previous study (**7**). In a few cases, *E* merely entered the room and made notes as *S* spoke. This procedure was used when only one aspect of the dream content was of interest, such as whether or not an outside stimulus had been incorporated. Special procedures will be described in the appropriate sections of the results.

RESULTS

A few general remarks concerning the dream recall obtained in this study might be made to facilitate understanding of the results. The incidence of recall after awakenings during rapid eye-movement periods was about the same, 80%, as has been previously reported (**7**). In order to gain maximum recall, *Ss* had to focus their attention on the dream content as soon as they were aware of being awake. An abrupt arousal seemed to aid this process, and if they were distracted from the task of remembering for even a few seconds, the dream was often forgotten.

The *Ss* were usually allowed to describe their dreams in their own way. This, of course, resulted in considerable variation in the narratives depending on *S*'s loquacity or propensity for detailed description, and on the relative importance attached to different elements of the dream. Thus, one *S* might merely say that he had been walking down a street, while another might give a

detailed account of the buildings and people on the street, the weather, and so forth. In general, it seemed that relatively neutral experiences tended to be glossed over while more exciting ones were described at great length. It is also likely that many dreams were more tersely described than they might have been because of *S*s' desire to get back to sleep as soon as possible. Questions on certain points after *S* had finished talking would invariably elicit more detail.

Dreams and Body Movements

It has been previously observed (**8**) that the cessation of eye-movement periods was often accompanied by a large body movement on the part of the sleeper, and that during the eye-movement periods the incidence of gross body movement was markedly lower than during the adjacent intervals. This suggested the possibility that body movements interrupting eye-movement periods, yet followed by additional ocular activity, might mark the end of one dream sequence and the beginning of another.

In order to test this possibility, the dream transcriptions were carefully screened for examples of either long narratives of smoothly related, continuously progressing events, or accounts consisting of two or more distinct, seemingly unrelated, dream fragments. Dreams that could unequivocally be placed in one of the two categories were relatively infrequent. Many were eliminated because of poor recording. Others did not obviously fit in either category. An example of the latter was when the dream contained an abrupt shift of scene, as from a restaurant to the home dining room, but with the same people remaining as central characters. Thus, it was neither smoothly continuous nor were the two scenes entirely unrelated.

A total of 46 dreams fulfilled the requirements of being long and continuous, and 31 contained two or more apparently unrelated fragments. The associated electrical recordings were then examined for the presence of artifactual signs of body movement and were separated into those in which no body movements occurred during the eye-movement period, and those containing signs of one or more body movements. A two-by-two contingency table (Table 1) was used in which the presence or absence of body movements was related to recall of continuous or fragmented dreams. From this analysis, it appears that long continuous dreams are more likely to be recalled in the absence of gross body movements while fragmented dreams are more likely to be recalled if the periods contain body activity.

Assuming the execution of a body movement actually marked the end of a discrete dream episode, it seemed possible that this fragment might be forgotten and the recall limited to the dream activity following the movement and preceding the awakening. With this in mind, the transcriptions were again scanned for narratives which were much shorter than would have been expected from the objective duration of the eye-movement period. In 45 instances, dream narratives after 10 to 20 min. of rapid eye movements were as short or shorter than those generally obtained after 4 or 5 min. of eye activity. It was found that in 37 of these 45 cases, a large body movement had occurred several minutes before the awakening.

As a result of these findings, it was decided to awaken *S*s during eye-move-

TABLE 1

RELATIONSHIP OF BODY MOVEMENTS TO DREAM CONTINUITY*

Body Movement	Dream Continuity				Total
	Fragmented		Continuous		
	Observed	(Expected)	Observed	(Expected)	
Present	21	(17.5)	14	(17.5)	35
Absent	10	(21)	32	(21)	42
Total	31		46		77

* 10.83 is the critical point when $P = .001$, $df = 1$. In this sample $\chi^2 = 12.9$, $P < .001$.

ment periods just *after* a gross body movement was observed. This was done 10 times. In 3, no content was recalled, in 3, the dream was reported as usual, but in the other 4, the following very suggestive remarks were spontaneously made by *S*s:

A. "I had just come to the end of a dream, I think, but it was a natural transition point. I had been talking to a friend and . . . [dream described]. Then I had the feeling that the dream ended before I woke up."

B. "I don't know if I was awake when I heard the bell. The dream seemed to be over before the bell rang."

C. "I'm trying to decide if I was dreaming just when you woke me up. I guess I was . . . [described dream]. I kind of had this feeling that just about as I finished the dream just mentioned there seemed to be a kind of little blank space or something and then just a picture of a telephone or something [arousing bell]. I got the idea that I'd finished before the bell sounded."

D. "[Dream described] I had the impression that the dream had reached its natural end, that the TV program I was watching had reached its fade-out point."

It is also worth mentioning that in one *S*, who was awakened during an eye-movement period that had a body movement every 2 or 3 min., the dream recall was extremely disjointed, consisting of six unrelated fragments. The *S* spontaneously commented that he couldn't understand why the dream was so "mixed up." All these findings indicate that body movements during dreams tend to signal a change in the dream activity.

Eye Movements and Dream Activity

The term "eye-movement period" used in this and in other communications, refers to a discrete interval of sleep during which rapid eye-movement potentials are seen in association with a characteristic low voltage, nonspindling, EEG pattern which persists unchanged throughout the entire period. Such periods are preceded and followed by intervals during which no rapid eye movements are seen and the EEG pattern is distinctly different, containing slow waves and sleep spindles. However, the degree of eye movement shows extreme variability during the period of ocular activity although the EEG pattern remains constant. Accordingly, it has been hypothesized that the variability in ocular activity depends on the specific activity of the dreamer within his dream, and therefore the physiological variables, amount and direction of eye movements, should allow prediction of the psychological variables, degree of involvement of the dreamer in the action of the dream and the spatial orientation of the dream activity, i.e., horizontal or vertical.

Amount of ocular activity as a measure of the degree of involvement of the dreamer in the dream.—This series of observations was made in the attempt to test the hypothesis that those dreams physiologically characterized by frequent and large eye movements would be marked by an active participation of the dreamer in the dream activity, while those dreams characterized by infrequent and relatively small eye movements would be distinguished by passive participation of the dreamer.

In order to test this hypothesis independent of possible bias on the part of *E*, a series of dreams originally obtained for an entirely different purpose was utilized. The electrical recordings and dream transcriptions were separated after matching codes had been assigned to each. The *E*s first characterized the dream transcripts as "active" or "passive," discarding those dreams where the transcription was incomplete or too ambiguous for agreement. Of the 180 dream transcripts available, only 109 could be categorized on the basis of the content as either active or passive. Had the dreams been originally obtained for this purpose, the diminution in amount of data due to ambiguity could have been avoided by careful questioning. For the purposes of this study, active dreaming was taken to include those dreams in which the dreamer was an active participator in the events of the dream, whereas passive dreaming was taken to include those dreams in which

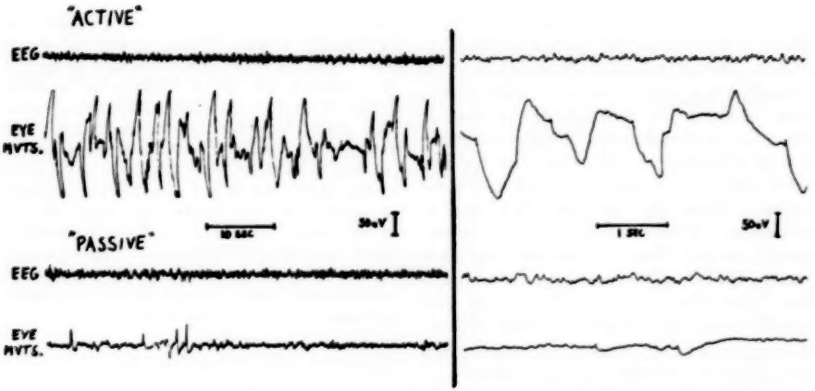

FIG. 2. Examples of "active" and "passive" eye-movement recordings. The active record contains many large movements, while in the passive record the movements are few and small. The EEG patterns are the same in both.

S was quietly reflecting upon an event, talking quietly to another person, or watching an event occur, often from a distance, in which he took no active part. The greatest loss of data involved dreams in which it was not clear from the transcript whether the dreamer was watching the event or participating in it. A good example of such a dream, where ambiguity would exist unless clearly described, is the watching of a television program. The dreamer might describe considerable action although he was actually a passive onlooker, and the entire television screen could have been included in a single fixation of the eyes.

The transcriptions were then set aside and the eye-movement records were independently classified as either active or passive on the basis of the eye-movement potentials. Figure 2 shows sample records of the two categories. Those records containing a large amount of eye movement were called active and other records containing minimal eye movement were classified as passive. Specific criteria were set up to distinguish borderline records, but, as a rule, the differentiation was quite obvious. Only four of the records associated with the 109 dreams could not be classified because of an excess of confusing artifacts. Because of the fact previously mentioned, that several distinct dream episodes were often associated with a single eye-movement period, only the last episode in the content was classified, and therefore, also, only the last 5–10 min. of eye-movement potential record before the awakening. The dream classifications were then compared to the eye-movement ratings by means of a two-way contingency table (Table 2). It appears from this analysis that eye-movement periods with minimal ocular activity are associated with passive dreams, while frequent and large eye movements are associated with active dreams.

TABLE 2

RELATIONSHIP OF *S*'s EYE ACTIVITY TO RECALLED DREAM CONTENT*

Eye Movement	Dream Content				
	Active		Passive		
	Observed	(Expected)	Observed	(Expected)	
Active	38	(29)	20	(29)	58
Passive	5	(23.5)	42	(23.5)	47
Total	43		62		105

* 10.83 is the critical point when $P = .001$, $df = 1$. In this sample $\chi^2 = 34.6$, $P < .00001$.

The direction of the eye movement as a direct expression of the spatial orientation of the dream activity.—Actually, it is hypothesized that the eye movements seen in sleep are those that the sleeper executes in the dream. However, since eye movements and fixations are for the most part not conscious, they must be inferred from the activity that one is watching. Thus, vertical eye movements should result when one watches events taking place in a vertical plane. For very short events, a person can usually remember where he "looked," although he does not remember consciously executing the specific eye movement.

In order to test the hypothesis that rapid eye movements are an exact indication of the direction of the dream activity, the very last action reported by *S*, i.e., where he looked, was compared with the very last eye movement (or its lack) before the awakening. Six *S*s were studied over 10 nights of sleep using both horizontal and vertical electrode placements around the eyes. Control fixations were executed while *S* was awake to determine the potential form associated with each of the four directions of eye movement. Of 39 dreams elicited in this series, the last dream activity before awakening could be recalled in 23 instances with sufficient clarity to differentiate the direction of the last eye movement in the dream. The last eye movement in the record was identical with the last reported fixation in the dream in 17 cases. In one example of this procedure, the last recorded eye movement was downward. The last reported dream activity was picking up an object on the floor. In this case, as in others, the eye movement corresponded in direction to what would have been expected if the dreamer had carried out this action in real life. Since *S*s had to choose the correct movement from among five possibilities (up, down, right, left, none), the probability of attaining the 74% accuracy by chance alone was vanishingly small.

In addition to the above findings, two anecdotal examples similar to the others previously reported in support of this hypothesis (**7**) have been obtained. The dreams were elicited after awakenings which followed an interval during which *only* vertical eye-movement potentials were seen in the record. (Instances in which either vertical or horizontal movement persists as a pure component are extremely rare.) In one, *S* was standing before a staircase leading up into a building. He was anxiously peering at the address which was carved in the sidewalk at his feet, and then looking up toward the top of the stairs. In the other, *S* was watching a blimp that hovered above him. The occupants of the blimp began dropping leaflets and *S* recalled alternately looking up at the blimp and down at the dropping leaflets.

Effect of Stimuli on Dreams

Since it is now possible to determine when *S* is dreaming by objective criteria, the effect of external stimuli can be more accurately assessed. In this study, three external stimuli were used: the sounding of a tone, flashing of a light, and a spray of cold water. These stimuli represented different modalities as well as different degrees of what might be called noxiousness, cold water being the most noxious of the three, and the pure tone the least.

The tone stimulus was a 1000-cycle pure tone sounded for 5 sec. The awakening threshold during eye-movement periods was previously determined and an intensity slightly below this threshold was used. In spite of this, the tone occasionally awakened the sleeper. The flashing of a 100-w. lamp placed where it would shine directly on the sleeper was the light stimulus. In these tests, it was necessary to enter the sleeping room to be certain that *S* was lying with his face exposed to the light. The light flashes never awakened *S*. The third stimulus was a fine spray of cold water ejected from a hypodermic syringe. In several instances, *E* moved close to the *S* and allowed several large drops of water to fall on an exposed part. This stimulus most frequently awakened *S*s. In general the stimuli were given after eye movements and the characteristic EEG had appeared, and then *S*s were awakened by the doorbell a few minutes later.

It was not always easy to decide

TABLE 3

INCORPORATION OF EXTERNAL STIMULI INTO REPORTED DREAMS

Stimulus	Tests Without Awakening	Incorporations Into Dream		Tests Resulting in Awakening	Modifications at Awakening	
		No.	%		No.	%
1000-cps tone	35	3	9	5	0	0
Light flashes	30	7	23	0	0	0
Water spray	33	14	42	15	6	40
Arousing bell	0	0	0	204	20	10
Total	98	24	24	224	26	12

whether or not a stimulus had modified the dream. If a light had been flashed in *S*'s eyes, one might get a subsequent dream containing light simply because it was daytime. Consequently, one could feel certain that the stimulus had been incorporated into the dream only if there was an obvious modification at the appropriate time. If the eye movements had been present for, say, 10 min., the stimulus then applied, and the sleep allowed to continue for another minute, one would expect the modification to occur toward the end of the complete dream sequence.

Table 3 summarizes the results of these experiments in which 12 *S*s participated. In addition to the above described stimuli, the available dream transcripts were analyzed to see how often the awakening doorbell had been incorporated. Examples of the types of modification seen were as follows:

A. *Tone.* The sudden intrusion of a brief, roaring sound into the dream. The dreamer was frightened and thought that there was either an earthquake or that a plane had crashed outside his house.

B. *Light.* The modifications included a sudden fire, a flash of lightning, seeing shooting stars, and a dream of *E* shining a flashlight into *S*'s eyes.

C. *Water spray.* Typical examples included sudden rainfalls, leaking roofs, and being squirted by someone.

D. *Awakening doorbell.* The most frequent incorporations were either the telephone or doorbell ringing just as *S* awoke.

None of the stimuli were markedly effective in modifying the dreams. The water spray was the most effective and the pure tone was the least, and it might be pointed out that the latter represented a relatively bland stimulus in contrast to the cold water.

The Ss were also stimulated 15 times during the interspersed periods of ocular quiescence. The stimuli were divided equally among the three types. In no case was an eye-movement period initiated by an external stimulus and no dreams were recalled when *S*s were awakened shortly after the stimuli were given.

An attempt was made to test the effect of an internal stimulus. Three *S*s on five occasions completely restricted their intake of fluids for periods of 24 hr. or more before sleeping in the laboratory. On all occasions they reported that they were very thirsty when they went to bed, and twice their thirst had reached distressing proportions with dry lips, inability to salivate, etc. Fifteen narratives were elicited under these conditions. In no case did the dream content involve an awareness of thirst or descriptions of actually drinking something. Ten dreams seemed completely unrelated to this internal stimulus, and five contained elements that might have been related to thirst as follows:

A. "I was in bed and was being experimented on. I was supposed to have malabsorption syndrome."

B. "I started to heat a great big can of—a great big skillet of milk. I put almost a quart of milk in."

C. "Just as the bell went off, somebody raised a glass and said something about a toast. I don't think I had a glass."

D. "While watching TV I saw a commercial. Two kids were asked what they wanted to drink and one kid started yelling, 'Coca-Cola, orange, Pepsi,' and everything."

E. "I was watching a TV program, and there was a cartoon on with the animals that are like those in the Hamm's beer advertisement."

The Course of Time in Dreaming

Because, in at least a few instances, stimuli were incorporated into the dream content without awakening *S*, it was possible to time short intervals of dreaming. The *S*s were stimulated during eye-movement periods by one of the methods described earlier, and after a precise interval of time they were awakened. In each of 10 instances, where the stimulus was incorporated and the subsequent interval was precisely timed, the amount of dream action in the interval between the modifying stimulus and the awakening did not vary far from the amount of action that would have been expected to take place during an identical time in reality. Several examples will suffice to illustrate this:

A. The *S* was sleeping on his stomach. His back was uncovered. An eye-movement period started and after it had persisted for 10 min., cold water was sprayed on his back. Exactly 30 sec. later he was awakened. The first part of the dream involved a rather complex description of acting in a play. Then, "I was walking behind the leading lady, when she suddenly collapsed and water was dripping on her. I ran over to her and felt water dripping on my back and head. The roof was leaking. I was very puzzled why she fell down and decided some plaster must have fallen on her. I looked up and there was a hole in the roof. I dragged her over to the side of the stage and began pulling the curtains. Just then I woke up."

B. The *S* was sleeping on his back. It was a hot summer night, and he was completely uncovered. An eye-movement period started, and after 5 min. cold water was sprayed on his feet and legs. One minute later he was awakened. The first part of the dream involved being in a room talking to some friends. Then, "two children came into the room and came over to me asking for water. I had a glass of ice water and I tipped the glass to give it to them. I was sitting, and I spilled the water on myself. The children wanted the ice and tried to grab it, but it slipped away. I got mad, because they were so greedy, and tried to shove them away from the chair. Then I got out of the chair and was going to change my pants. As I left the room I seemed to be in a school and I saw many children in the hall and I seemed to be late for a class. I was starting to climb up some stairs when the bell rang."

C. A third example is included because of its special interest. It was obtained fortuitously. The *S* was standing by a record player in his dream listening to some music, intending to leave for home. "The doorbell rang and she asked me if I would answer it. I hesitated for a moment and, as I started to go, it rang again." In this instance *E*'s finger accidently slipped off the awakening bell, and hence it was rung twice to rouse *S*. About 3 or 4 sec. elapsed between the two rings.

Discussion

It is of historical interest that the possible relation of eye movements to dreaming was suggested in 1892 by George Trumball Ladd in an essay on visual dreams (**12**). This startlingly accurate hunch was not subjected to experimental testing, however, and it was many years later that Aserinsky and Kleitman actually demonstrated such a relationship. The results of the present study indicate that the eye movements are not only related to dreaming, but are a specific reproduction of the dreamer's visual activity, and that the examination of these eye movements enables prediction of certain elements of the dream content.

Gross body movements appear to be associated mainly with the termination of a dream period and with the end of dream sequences within the period. It cannot be stated definitely whether a body movement simply interrupts the dream, or whether the dream sequence reaches a natural termination and with it some inhibitory influence also ceases, thereby allowing a body movement to take place. The fact that the incidence of body movement was found to be higher in the intervals immediately preceding and following eye-movement

periods (**8**) suggests that the latter interpretation is more accurate. Also *Ss* occasionally reported that the dream seemed to have reached a natural conclusion if they were awakened immediately after a body movement had occurred.

It seems somewhat paradoxical that, whereas the eye movements seem to be specifically related to the dream imagery, body movement does not. Although part of the objective stillness of the dreamer can be explained by assuming he is held fascinated by the drama of the dream, it is impossible to account for dreams in which the dreamer is running, fighting, etc. Since the eye movements so closely approximate movements carried out in the dream, one would expect that running or fighting dreams would be associated with at least rudimentary running or punching movements of the limbs. The only explanation one can offer at this point is that eye movements are apparently controlled from different cortical areas than most other movements (**10**), and for some reason the areas subserving the latter are not active during dreaming.

An old theory of dream causation is that dreams are initiated by outside stimuli. This notion seems untenable in view of the fact that when stimuli were given when the eyes were quiescent, neither eye movements nor dream recall resulted. Also, eye-movement periods occur regularly in the absence of any deliberate stimulation and entirely independent of chance stimuli such as noises outside the building, doors closing, and so forth.

Even if external stimuli do not initiate dreaming, the question of their importance in modifying the dream content must still be considered. In 1893, Calkins (**3**) reported a careful analysis of 375 dreams obtained by introspection from two *Ss* and found that only 33 contained elements that were traceable to actual physical sensations occurring during sleep. Weed and Hallam (**15**), repeating Calkins study, found only 20 of 381 dreams to have been influenced by external stimuli. On the other hand, Cubberly (**5**), experimenting with constant tactile sensation applied by means of gummed paper to a restricted portion of the body during sleep, reported that in 95% of the dreams an effect could be seen either in the form of a general tenseness of the dreamer or as a tension of the particular part of the body on which the gummed paper had been placed. However, since Cubberly apparently was his own *S*, there is a strong likelihood that autosuggestion played a part in his results. The findings reported in the present study indicate that external stimuli are not markedly effective in modifying the dream content, but that some (cold water) are more effective than others (pure tone). Although external stimuli can at times interject a new element, the basis for the images and particular story of a dream must exist mainly within the dreamer.

Considerable differences of opinion exist regarding the course of time in the dream. The idea that a dream which appears to last many minutes may actually last but a few seconds is based mainly on a single type of evidence. Occasionally, dreams have been reported in which a lengthy series of events was logically related to an awakening stimulus. One of the most famous examples is a dream reported by Maury (**14**) in which he was in Paris during the Reign of Terror. After witnessing a number of frightful scenes of murder, he was finally himself brought before the revolutionary tribunal. There he saw Robespierre, Marat, and other prominent figures of those terrible days. He was questioned, and condemned, and led to the place of execution surrounded by an immense mob. He climbed onto the scaffold and was bound to the plank by the executioner. The blade of the guillotine fell. He felt his head being separated from his body, woke up in extreme anxiety—and found that the top of the bed had fallen down and had struck his cervical vertebrae just in the way in which the blade of the guillotine would actually have struck. Since in this and other examples, the lengthy

events were so appropriate to the awakening stimulus, it has been proposed that the stimulus initiates the dream and all the imagery is compressed into the short interval between the initial perception of the stimulus and the awakening. Considering the very large amount of dream literature, these examples are quite rare, and pertaining to Maury's dream, Ellis (**9**) has pointed out that the dream was not recorded until ten years after it had been dreamed.

Anecdotal reports have also lent support to the contrary view that the duration of a dream is identical to the duration of a similar experience in reality. Clavière (**4**) has reported a dream in which the sound of an alarm was incorporated twice into a dream. In this dream a conversation occurs between the two alarms which, when Clavière repeated it in waking life, took 17–18 sec. Upon testing the alarm clock, he found that 22 sec. elapsed between the time of the first and second rings. A similar anecdote was reported in our results with a like correspondence between dream and reality.

In an earlier study of eye movements and dream recall, it was found that the length of eye-movement periods correlated with the length of dream narratives. Further, the action reported in the narratives corresponding to different lengths of eye-movement periods was such that the same amount of time might have been expected to elapse if the dreams were re-enacted. Such findings support the hypothesis that dream events and real events proceed at about the same rate. The results of the attempts to time dreams in this study also support this idea. An additional corroboration is that the speed and duration of the eye movements themselves is the same during sleep as it is in wakefulness.

One implication of the apparent fact that one can predict certain elements of the dream from physiological measurements is that it is almost certain in many cases that the dreamer reported the dream as it actually occurred. It has been suggested that the dream report, even if sincere, is a disguised, secondary elaboration of the "real" dream. If this were true, it seems very unlikely that the objective eye-movement patterns would correspond so well to the dream content. Of course, this may only apply to dreams elicited in the laboratory where the verbal report is obtained immediately after an awakening that interrupts an eye-movement period. If the dream recall is obtained at a later time, either upon awakening in the morning or during the following day, it is possible that the secondary elaboration postulated by others may occur between the actual time of dreaming and the time of the delayed report.

From this and other studies it can be seen that certain aspects of dreaming which in the past could only be speculated about may now be subjected to experimental analysis. Using eye-movement potentials and brain waves as objective criteria of dreaming, it has been possible to determine the pattern of occurrence, duration, and course of time of dreams, as well as the relation of specific eye movements, body movements, and external stimuli to their content. It is further hoped that this technique may afford a means of assessing certain theories of the meaning of dreams and perhaps give some insight into their fundamental significance to the physiological and psychological economy of the organism.

Summary

The occurrence of dreaming was localized in 16 *Ss* by observation of rapid eye movements and characteristic brain-wave patterns. The dream content recalled after awakening at these times was subjected to analysis. It was found that the amount of observed eye movement was related to the degree of participation of *Ss* in the events of the dreams. The direction of the last observed eye movement before awakening corresponded to the last reported fixation of the dreamer. Body movements seemed to be related to the termination of a dream sequence. Certain external and internal stimuli were not found to be important factors influencing the course of the dream content. By timing the interval between the introduction of a modifying stimulus and the subsequent awakening, it was found that the course of time in the dream was

comparable to the time elapsing during an identical activity performed in the waking state. The implications of these findings was discussed.

References

1. Aserinsky, E., & Kleitman, N. Two types of ocular motility occurring in sleep. *J. appl. Physiol.*, 1955, **8,** 1–10.
2. Bergman, P. S., Nathanson, M., & Bender, M. B. Electrical recordings of normal and abnormal eye movements modified by drugs. *Arch Neurol. Psychiat.*, 1952, **67,** 357–374.
3. Calkins, M. W. Statistics of dreams. *Amer. J. Psychol.*, 1893, **5,** 311–343.
4. Clavière, J. La rapidité de la pensée dans le rêve. *Rev. Phil.*, 1897, **43,** 507–509.
5. Cubberly, A. J. The effect of tensions of body surface upon normal dream. *Brit. J. Psychol.*, 1923, **13,** 243–265.
6. Dement, W. Dream recall and eye movements during sleep in schizophrenics and normals. *J. nerv. ment. Dis.*, 1955, **122,** 263–269.
7. Dement, W., & Kleitman, N. The relation of eye movements during sleep to dream activity: an objective method for the study of dreaming. *J. exp. Psychol.*, 1957, **53,** 339–346.
8. Dement, W., & Kleitman, N. Cyclic variations in EEG during sleep and their relation to eye movements, body motility, and dreaming. EEG clin. Neurophysiol., 1957, **9,** 673–690.
9. Ellis, H. The world of dreams. New York: Houghton Mifflin, 1922.
10. Fulton, J. F. Physiology of the nervous system. (3rd ed.) New York: Oxford Univer. Press, 1949.
11. Kris, C. Diurnal variation in periorbitally measured eye potential lev l. *EEG clin. Neurophysiol.*, 1957, **9,** 382.
12. Ladd, G. T. Contribution to the psychology of visual dreams. *Mind,* 1892, **1,** 299–304.
13. Lion, K. S., & Brockhurst, R. J. Study of ocular movements under stress. *Arch. Ophth.*, 1951, **46,** 315–318.
14. Maury, A. Le sommeil et les rêves. Paris: Didier, 1861.
15. Weed, S., & Hallam, F. A study of dream consciousness. *Amer. J. Psychol.*, 1896, **7,** 405–411.

(Received June 19, 1957)

Journal of Experimental Psychology
Vol. 55, No. 6, 1958

ASSOCIATIVE INTERFERENCE IN VERBAL PAIRED-ASSOCIATE LEARNING[1]

NORMA F. BESCH

Columbus Psychiatric Institute and Hospital, Ohio State University

AND

WILLIAM F. REYNOLDS[2]

Western Michigan University

Associative interference has been designated as the negative transfer effect produced by the learning of one task upon the subsequent learning of another task. For verbal paired-associate learning, the A-B, A-C paradigm, in which the stimuli are identical in both lists but the responses are different, has been regarded as the optimum condition for producing negative transfer (**6**). First task responses are expected to reoccur on the transfer task in place of the new correct responses. However, Underwood (**7**), Atwater (**1**), and Morgan and Underwood (**4**) have found that interference was exhibited only in the first few trials of the transfer task by the A-B, A-C groups when they were contrasted with a control group which had received no prior learning. In fact, facilitation was observed for these groups later in learning.

An A-Br task is constructed by re-pairing first-list stimulus and response terms. Porter and Duncan (**5**) compared the performance of two groups given, respectively, A-Br and A-C transfer tasks. With the original A-B learning defining the control condition, the A-C and A-Br groups exhibited positive and negative transfer effects, respectively.

None of these investigators has utilized a control condition taking account of such factors as "warm-up" or "learning to learn"; therefore, facilitation, rather than interference, in the A-B, A-C paradigm has not been clearly demonstrated. However, Porter and Duncan did find substantial negative transfer for an A-Br task, despite the use of a control condition which would tend to obscure such an effect. In a more recent experiment which controlled for the occurrence of nonspecific transfer, Mandler and Heinemann (**2**) report positive transfer for A-Br pairs and no effect for A-C pairs. While familiar words were employed for the paired-associate material in the investigations previously cited, Mandler and Heinemann used consonant nonsense syllables for the response terms of their pairs. It is not surprising that Mandler and Heinemann failed to confirm previous findings since, in their situation, considerable learning of the response term itself would be required in addition to pair learning.

The present study is essentially a replication of the Porter and Duncan experiment. A within-*Ss* comparison was used in preference to a between-*Ss* comparison in order to provide a

[1] The authors wish to thank Kathryn J. Norcross for performing the statistical analysis. They would also like to express their appreciation to Charles C. Spiker for his aid in the design of the experiment and for his many helpful suggestions.

[2] This research was conducted while the authors were graduate students at the State University of Iowa. The experiment was conducted in the psychology laboratory of K. W. Spence.

control condition subject to the same nonspecific transfer effects as the experimental conditions.

Method

Subjects.—The *Ss* were 33 female volunteers from the introductory psychology course at the State University of Iowa. They were assigned at random to one of three groups. The three groups learned different forms of List I, which were constructed to control for any differential difficulty of transfer-list pairs.

Design.—All *Ss* learned two lists. List II, the transfer task, which was identical for all *Ss*, may be considered as consisting of three sets of four word pairs each. The three forms of List I were constructed to permit each of these three sets to serve in turn under each of the transfer-task conditions. The transfer-task conditions were: (*a*) A-Br (the stimulus and response terms from four of the List I pairs were re-paired), (*b*) A-C (four of the stimuli from the first list were paired with new response words), and (*c*) D-C (four new stimuli were paired with new response words).

Lists and apparatus.—The transfer task and the three forms of List I each consisted of 12 pairs of two-syllable adjectives, which were obtained from the Melton-Saffier list (**3**). The adjectives for the transfer task were chosen from the list to minimize similarity of meaning between paired stimulus and response terms. For the three forms of List I, only those pairs which were involved in the A-B, D-C condition permitted a selection on the same basis. Since few pairs with low-scaled values of similarity of meaning were available from the Melton-Saffier list, some intralist structural similarities were unavoidably present.

Two Hull-type memory drums were employed to present the lists of paired-associate material. The original task was presented on one drum; the transfer task was presented on the other. The stimulus items were exposed every 4 sec., including a 2.2-sec. anticipation interval. A 4-sec. rest interval was given between successive presentations of a list. Three presentation orders were used for each list to prevent serial learning.

Procedure.—Standard verbal-learning instructions were given before List I; prior to the presentation of the transfer task, *Ss* were informed that the same procedure was to be followed in the learning of a new list. The time interval between the tasks was approximately 3 min. Both lists were learned by the anticipation method to a criterion of one errorless trial.

TABLE 1

Mean Number of Correct Responses to Criterion on the Transfer Task

Groups	Pair Condition		
	D-C	A-C	A-Br
I	69.27	53.18	48.73
II	58.00	55.91	49.45
III	53.36	40.36	40.45
Total	60.21	49.82	46.21

TABLE 2

Analysis of Variance of the Total Number of Correct Responses to Criterion on the Transfer Task

Source	*df*	*MS*	*F*
Total between *Ss*	32		
Groups (B)	2	1394.37	1.26
Error	30	1106.91	
Total within *Ss*	66		
Pair conditions (A)	2	1743.70	30.06*
A × B	4	187.44	3.23**
Residual	60	58.00	

* $P < .001$.
** $P < .025$.

Results

Table 1 presents the mean number of correct responses to criterion on the transfer task for each pair condition. The analysis of variance, summarized in Table 2, yielded an F of 30.06 ($P < .001$) for the main effect of Pair conditions, and an F of 3.23 ($P < .025$) for the interaction be-

TABLE 3

t Tests of Differences Between Pair-Conditions for Each Group

Comparisons	Group I	Group II	Group III
A-Br & D-C	10.93**	4.55**	6.87**
A-C & D-C	8.56**	1.11	6.92**
A-Br & A-C	2.37*	3.44**	.05

** $P < .01$.
* $P < .02$.

tween Groups and Pair conditions. In view of the significant interaction, t tests of the differences between the various pair conditions were performed for each group separately. An inspection of Table 3 reveals that, with the exception of two comparisons, all the differences were significant. Thus, it may be concluded tentatively, that while both the A-Br and A-C pairs are subject to interference, the A-Br pairs are affected more.

The significant interaction appears to be attributable to the lack of a difference between the performances on the A-C and A-Br pairs for Group III. It may be relevant that the mean number of trials which were required to reach criterion on Task I was 22.00 (SD = 7.21) and 25.55 (SD = 6.80) for Groups I and II, respectively, while Group III required only 18.55 (SD = 5.16) trials. However, a correlational analysis failed to support the existence of a relationship between rate of learning on List I and the difference between A-Br and A-C pair learning. It should be noted that an F of 3.23 ($.05 < P < .10$) for Groups was obtained from a simple analysis of variance of trials to criterion on Task I.

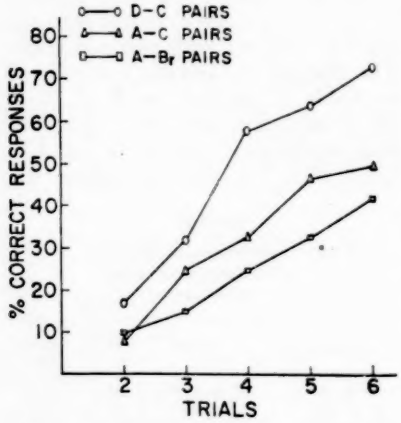

FIG. 1. Percentage of correct responses for pair conditions in Trials 2–6 of transfer task.

TABLE 4

SOURCE AND NUMBER OF INTRUSIONS DURING TRANSFER TASK LEARNING

Pair-Source of Intrusions	Pair Condition		
	D-C	A-C	A-Br
A-Br	10	57	108
Other	47	39	80
Total	57	96	188

It is apparent from an inspection of Fig. 1 that the effects exhibited in the initial trials of the transfer task are consistent with the over-all transfer effects. The analysis of variance of total correct responses on Trials 2–6 yielded F's of 21.44 ($P < .001$) and 118.38 ($P < .001$) for Pair conditions and Trials, respectively. No interaction was present, unlike the previous analysis, between Groups and Pair conditions.

Table 4 shows that the frequency of intruding responses, defined here as an incorrect anticipation, exhibits a trend for pair conditions similar to the one found for correct responses. It is also clear that the majority of the intrusions for both the A-Br and A-C pairs are the response terms appropriate to the A-Br pair set. In addition, it was observed that intruding responses on List II which were correct responses on List I occurred 83 times for the A-Br pairs and three times for the A-C pairs.

DISCUSSION

One interpretation to account for the disparity of the present results with respect to previous findings may be

offered in terms of the generalization of habit or associative strength during transfer-task learning. It will be recalled that Porter and Duncan (5) demonstrated interference for A-Br pairs, and they found, consistent with other investigators, facilitation for A-C pairs. The primary feature which distinguishes the present study from the Porter and Duncan experiment is that the A-C pairs were learned in the same list as the A-Br pairs rather than in a separate list. Thus, the stimuli of the A-C as well as the A-Br pairs are given an opportunity on the transfer task to establish generalized habit strengths to responses which were also present in List I, that is, the response terms of the A-Br pairs. For the A-Br pairs, the further development of associative strength by generalization between the stimulus and response items which were previously paired, maintains for some time tendencies of sufficient strength to compete successfully with the correct responses on the transfer task. For the A-C pairs, the further establishment of generalized habit strengths during the transfer task between the stimuli of these pairs and the response terms of the A-Br pairs yields competing tendencies with sufficient strength to cause interference.

It is implicit in the preceding statements that there are two habit components contributing to the total associative strength present for a competing tendency—one component from original learning and one from transfer task learning. Since the generalized habits developed during the transfer task are not expected to differ for the various pair conditions, the associations established during original learning would determine the relative strengths of the competing tendencies present for the different pair conditions. During original learning, associations are established between paired stimulus and response terms and by generalization between nonpaired items. Presumably the connections established between paired items would be of greater strength than those between nonpaired items. The stimulus and response items which constitute the competing tendencies for the A-Br pairs were paired on List I, while the items of the A-C pairs were nonpaired. From a consideration of the competing tendencies alone, it would be predicted that the A-Br pairs would be subject to more interference than the A-C pairs. Clearly, the least interference would be shown in the case of the D-C pairs, since their stimuli were not present during original learning, thereby precluding the development of habit strength from that source. The fact that the correct tendencies for the A-C and D-C pairs have little or no habit loading at the beginning of transfer-task learning, while the items of the correct tendencies for the A-Br pairs were subject to generalization effects during original learning, limits the generality of the predictions.

Summary

Thirty-three *Ss* learned successively two lists of paired two-syllable adjectives. List II consisted of three kinds of pairs: (*a*) A-Br (stimulus and response terms from the first list were re-paired), (*b*) A-C (new responses were paired with stimuli from List I), and (*c*) D-C (new stimulus and response terms were paired).

In comparison to D-C pair learning, the A-Br and A-C pairs exhibited significant amounts of negative transfer in five of the six comparisons. Significantly fewer correct responses were made for the A-Br pairs than for the A-C pairs in two of the three comparisons. An inspection of the intrusion data revealed that the intruding responses for both the A-Br and A-C pairs were primarily those appropriate to the A-Br pair set.

The results were interpreted in terms of the competitive action of stimulus-response tendencies whose associative strengths were developed during both original and transfer-task learning.

References

1. Atwater, S. K. Proactive inhibition and associative facilitation as affected by degree of prior learning. *J. exp. Psychol.*, 1953, **46**, 400–404.

tween Groups and Pair conditions. In view of the significant interaction, t tests of the differences between the various pair conditions were performed for each group separately. An inspection of Table 3 reveals that, with the exception of two comparisons, all the differences were significant. Thus, it may be concluded tentatively, that while both the A-Br and A-C pairs are subject to interference, the A-Br pairs are affected more.

The significant interaction appears to be attributable to the lack of a difference between the performances on the A-C and A-Br pairs for Group III. It may be relevant that the mean number of trials which were required to reach criterion on Task I was 22.00 (SD = 7.21) and 25.55 (SD = 6.80) for Groups I and II, respectively, while Group III required only 18.55 (SD = 5.16) trials. However, a correlational analysis failed to support the existence of a relationship between rate of learning on List I and the difference between A-Br and A-C pair learning. It should be noted that an F of 3.23 ($.05 < P < .10$) for Groups was obtained from a simple analysis of variance of trials to criterion on Task I.

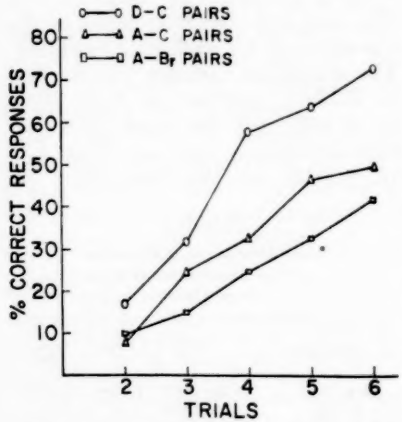

FIG. 1. Percentage of correct responses for pair conditions in Trials 2–6 of transfer task.

TABLE 4

SOURCE AND NUMBER OF INTRUSIONS DURING TRANSFER TASK LEARNING

Pair-Source of Intrusions	Pair Condition		
	D-C	A-C	A-Br
A-Br	10	57	108
Other	47	39	80
Total	57	96	188

It is apparent from an inspection of Fig. 1 that the effects exhibited in the initial trials of the transfer task are consistent with the over-all transfer effects. The analysis of variance of total correct responses on Trials 2–6 yielded F's of 21.44 ($P < .001$) and 118.38 ($P < .001$) for Pair conditions and Trials, respectively. No interaction was present, unlike the previous analysis, between Groups and Pair conditions.

Table 4 shows that the frequency of intruding responses, defined here as an incorrect anticipation, exhibits a trend for pair conditions similar to the one found for correct responses. It is also clear that the majority of the intrusions for both the A-Br and A-C pairs are the response terms appropriate to the A-Br pair set. In addition, it was observed that intruding responses on List II which were correct responses on List I occurred 83 times for the A-Br pairs and three times for the A-C pairs.

DISCUSSION

One interpretation to account for the disparity of the present results with respect to previous findings may be

offered in terms of the generalization of habit or associative strength during transfer-task learning. It will be recalled that Porter and Duncan (5) demonstrated interference for A-Br pairs, and they found, consistent with other investigators, facilitation for A-C pairs. The primary feature which distinguishes the present study from the Porter and Duncan experiment is that the A-C pairs were learned in the same list as the A-Br pairs rather than in a separate list. Thus, the stimuli of the A-C as well as the A-Br pairs are given an opportunity on the transfer task to establish generalized habit strengths to responses which were also present in List I, that is, the response terms of the A-Br pairs. For the A-Br pairs, the further development of associative strength by generalization between the stimulus and response items which were previously paired, maintains for some time tendencies of sufficient strength to compete successfully with the correct responses on the transfer task. For the A-C pairs, the further establishment of generalized habit strengths during the transfer task between the stimuli of these pairs and the response terms of the A-Br pairs yields competing tendencies with sufficient strength to cause interference.

It is implicit in the preceding statements that there are two habit components contributing to the total associative strength present for a competing tendency—one component from original learning and one from transfer task learning. Since the generalized habits developed during the transfer task are not expected to differ for the various pair conditions, the associations established during original learning would determine the relative strengths of the competing tendencies present for the different pair conditions. During original learning, associations are established between paired stimulus and response terms and by generalization between nonpaired items. Presumably the connections established between paired items would be of greater strength than those between nonpaired items. The stimulus and response items which constitute the competing tendencies for the A-Br pairs were paired on List I, while the items of the A-C pairs were nonpaired. From a consideration of the competing tendencies alone, it would be predicted that the A-Br pairs would be subject to more interference than the A-C pairs. Clearly, the least interference would be shown in the case of the D-C pairs, since their stimuli were not present during original learning, thereby precluding the development of habit strength from that source. The fact that the correct tendencies for the A-C and D-C pairs have little or no habit loading at the beginning of transfer-task learning, while the items of the correct tendencies for the A-Br pairs were subject to generalization effects during original learning, limits the generality of the predictions.

Summary

Thirty-three *Ss* learned successively two lists of paired two-syllable adjectives. List II consisted of three kinds of pairs: (*a*) A-Br (stimulus and response terms from the first list were re-paired), (*b*) A-C (new responses were paired with stimuli from List I), and (*c*) D-C (new stimulus and response terms were paired).

In comparison to D-C pair learning, the A-Br and A-C pairs exhibited significant amounts of negative transfer in five of the six comparisons. Significantly fewer correct responses were made for the A-Br pairs than for the A-C pairs in two of the three comparisons. An inspection of the intrusion data revealed that the intruding responses for both the A-Br and A-C pairs were primarily those appropriate to the A-Br pair set.

The results were interpreted in terms of the competitive action of stimulus-response tendencies whose associative strengths were developed during both original and transfer-task learning.

References

1. Atwater, S. K. Proactive inhibition and associative facilitation as affected by degree of prior learning. *J. exp. Psychol.*, 1953, **46**, 400–404.

2. MANDLER, G., & HEINEMANN, S. H. Effect of overlearning of a verbal response on transfer of training. *J. exp Psychol.*, 1956, **52,** 39–46.
3. MELTON, A. W. Materials for use in experimental studies of the learning and retention of verbal habits. Univer. of Missouri, 1936. (Mimeo.)
4. MORGAN, R. L., & UNDERWOOD, B. J. Proactive inhibition as a function of response similarity. *J. exp. Psychol.*, 1950, **40,** 592–604.
5. PORTER, L. A., & DUNCAN, C. P. Negative transfer in verbal learning. *J. exp. Psychol.*, 1953, **46,** 61–64.
6. UNDERWOOD, B. J. *Experimental psychology.* New York: Appleton-Century-Crofts, 1949.
7. UNDERWOOD, B. J. Proactive inhibition as a function of time and degree of prior learning. *J. exp. Psychol.*, 1949, **39,** 24–34.

(Received June 28, 1957)

Journal of Experimental Psychology
Vol. 55, No. 6, 1958

EFFECTS OF STRESS AND ANXIETY ON PERFORMANCE OF A COMPLEX VERBAL-CODING TASK [1]

LEON T. KATCHMAR, SHERMAN ROSS, AND T. G. ANDREWS

University of Maryland

The effects of psychological stress on performance have become an important area of research, as shown by the many independent studies recently published and by a number of reviews (**6, 9, 12, 17**). In this area, a number of studies have shown decrements in performance as a function of stress (**1, 3, 4, 11, 17**). Other studies have found no performance decrement under stress (**7, 24**), while still others have found both increments and decrements in performance under stressful conditions (**7, 13, 24**). The lack of uniformity of results may be attributed in part to the different dependent and independent variables used, and to differences in the motivation of the experimental *S*s used. The problem of motivational status is of special importance in studies dealing with psychological stress.

A number of studies have attempted to account for some of the differential results found in performance under stressful conditions by seeking relationships between various personality variables and performance. Such attempts have dealt with studies of manifest anxiety and rigidity (**2, 5, 20, 22**).

The present study deals with the effects of ego-involvement, falsified knowledge of results, and manifest anxiety on the flexibility of performance on a complex verbal-coding task. The type of behavior of interest here has often been discussed under the term "rigidity." Since there are theoretical connotations associated with this term, the term "flexibility" will be used. Flexibility refers to the ability of *S* to shift his performance in terms of the changing requirements of the situation. An additional feature of this study was an attempt to determine the effectiveness of the experimental variables by means of a stress experience inventory.

[1] This experiment is one of a series of studies on stress and behavior decrement performed under Contract No. DA-49-007-MD-222 between the Office of the Surgeon General, Department of the Army, and the University of Maryland.

METHOD

Design.—A basic 3 × 3 × 2 factorial design was used with three levels of failure stress, three levels of ego involvement, and two levels of anxiety, with 3 *S*s per cell. This design was used in the manner of Johnson and Tsao (**8**) which permits evaluation of the relative significance of individual differences as a separate source of variation in the experimental results.

Task.—The task employed in this experiment was the Woodworth-Wells Form Naming and Substitution Test (**23**). The test consists of 20 of each of the following five stimulus forms: star, circle, square, cross, and triangle. The 100 items were printed on a 9 × 9-in. card, arranged in 10 columns with 10 items to each column in a random order. The *S*'s task was to learn a code number for each of the five stimulus forms, e.g., *S* was required to learn to associate the number "1" with the star, "2" with the circle, "3" with the square, "4" with the cross, and "5" with the triangle. After learning to associate the numbers with their appropriate forms, *S*'s task was to read the entire list of 100 forms, alternately naming the stimulus forms *and* substituting the number code. The task, while apparently simple, provides a complex task which involves speed, accuracy, and flexibility. The task also provides a great deal of interference due to the necessity of making two different responses to the same stimulus.

Failure stress was imposed by evaluating *S*'s performance in comparison to *false* norms for

college and noncollege populations. Three levels of stress were used. The *Ss* in the no-stress group were told that their performance placed them at the 70th and 80th percentile, respectively, for the two populations, and this standing was about as we expected. The *Ss* in the medium-stress group were told that their performance placed them at the 40th and 50th percentiles, respectively, and that this standing was a little below the average of what was expected. The *Ss* in the high-stress group were told that their performance placed them at the 25th and 35th percentiles respectively and that this was considered very poor. Verbal and pictorial interpretations of the percentile norms appropriate to each of the three groups were presented to the *Ss*.

Ego involvement was manipulated by three different sets of instructions which defined the relative importance of the test to *S*. The first set defined the test as one which was interesting but of little relative importance. The second set defined the test as one which was used to assess intelligence. The third set defined the test results as a good indicator of success in college, and that the results might be used as a possible screening device.

The modified form of the Taylor Manifest Anxiety Scale (**20**) was administered to 257 students in introductory psychology courses at the University of Maryland under the title of a Biographical Inventory. The distribution was similar to that found by Taylor on a similar population.

Subjects.—Twenty-seven students scoring 24 and above and 27 making scores of 7 and below on the MAS were selected for the high- and low-anxiety groups, respectively. Each of these groups was composed of 18 men and 9 women students. Within each anxiety group, *Ss* were assigned at random to the various cells of the factorial design. In no case did more than one woman appear in any one cell.

Procedure.—Each *S* was tested individually in a relatively quiet, well-lighted room. Following the appropriate ego-involvement instructions, *S* was given a practice session consisting of approximately 10 min. of learning the items and the code, in addition to six trials on the test proper. The *Ss* were instructed that the test was a speed test, but that accuracy was also important. After the sixth practice trial, the appropriate stress instructions were given, followed by three more trials. The last three practice trials were actually a part of the test and were used to evaluate the change in performance under stress conditions.

After the last test trial, *S*, was given a Stress Experience Inventory to complete. This inventory consisted of 24 multiple-choice items. The inventory was subscored to provide estimates of anxiety, failure stress, and ego involvement. Four expert judges rated the 24 items according to whether they measured anxiety, failure stress, or ego involvement. The *S*'s responses to these related items constituted his poststress scores for each of the variables measured. At the completion of the inventory *E* described the nature of the experiment to remove any possible effects of the failure-stress condition.

Results and Discussion

The performance of each *S* was scored in terms of difference scores, errors, and frequency of blocking. A difference score was the difference between the average time in seconds to complete the last three practice trials and the three poststress trials. Errors and blocking were scored for the three poststress trials only. A "block" was defined as that elapsed time during which at least two responses could have been made. This score may more properly be considered a blocking score, reflecting the frequency of blocking.

The results of the analysis of the three performance measures along with the appropriate means for the major conditions are presented in Tables 1 and 2. For mean difference scores, a negative mean indicates a decrease in average time on poststress

TABLE 1

Means of Difference, Error, and Blocking Scores

Source	Level	Difference Scores	Error Scores	Blocking Scores
Anxiety	High	4.22	3.34	3.07
	Low	−5.44	2.19	1.86
Failure	None	−5.27	2.10	1.60
	Low	1.00	2.99	2.68
	High	2.83	3.20	3.11
Ego Involvement	Low	−3.00	2.85	2.23
	Medium	−1.17	2.51	2.31
	High	1.35	2.94	2.84

TABLE 2

ANALYSIS OF VARIANCE FOR DIFFERENCE, ERROR, AND BLOCKING SCORES

Source	Difference Scores		Error Scores		Blocking Scores	
	df	*MS*	*df*	*MS*	*df*	*MS*
Anxiety (A)	1	1251.85***	1	178537.01***	1	197653.13***
Failure (F)	2	355.02***	2	60865.16**	2	107694.50***
Ego Inv. (E)	2	62.57**	2	9046.56	2	20071.10
S's	2	637.91***	2	372761.18***	2	263622.68***
A × F	2	99.57*				
A × E			2	35610.65*		
F × E	4	55.49*				
A × I	2	.91	2	26846.56*		
A × F × I	4	1.29				
Error†	34	14.08	42	7989.60	46	9153.89
Total	53		53		53	

* $P = .05$.
** $P = .01$.
*** $P = .001$.
† Pooled insignificant interactions.

trials, while a positive mean indicates an increase in average time. As seen in Table 2, the variations attributable to the anxiety and failure-stress conditions for each of the three measures were highly significant. From the tabulated means, it appears that failure stress bears an inverse relationship to performance. The no-stress group decreased their average time following the stress instructions, in addition to making fewer errors and blocks on the poststress trials.

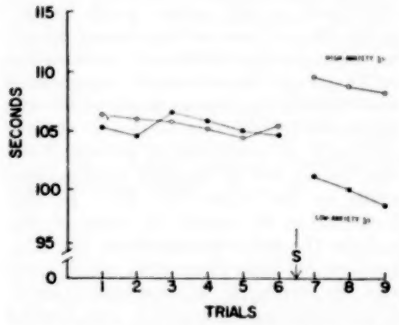

FIG. 1. Mean time to complete the task for high- and low-anxiety groups on pre- and post-stress trials. The stress instructions were given after Trial 6.

The medium- and high-stress groups, on the other hand, increased their average time following the imposition of stress. These groups also committed a greater number of errors, and frequency of blocking increased on the poststress trials. This change in performance was greatest for the high-stress group.

In regard to the anxiety classification, it appears that high-anxiety *S*s were affected by stress to a greater degree than the low-anxiety group. Figure 1 shows the average time taken by each anxiety group to complete each of the six pre- and three poststress trials. The two groups did not differ significantly on the prestress trials. Following the imposition of stress, however, the change in performance is marked. From the curves it appears that anxiety operates to sensitize an individual to the stress situation.

The effects of ego-involvement instructions on the average time taken to perform the test on pre- and post-stress trials are shown in Fig. 2. The three groups are clearly differentiable on the prestress trials. This

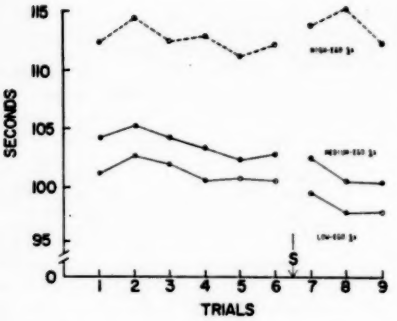

FIG. 2. Mean time to complete the task for each of the three ego-involvement groups on pre- and poststress trials. The stress instructions were given after Trial 6.

difference was significant at the .01 level. When prestress performance was partialed out by an analysis of covariance, the poststress differences were found to remain significant at the .05 level. This result is consistent with the analysis of difference scores presented in Tables 1 and 2. Error and block measures were not found to be sensitive to the ego-involvement conditions. The highly significant individual differences term found in each of the analyses clearly indicates the wide individual differences in performance of the *Ss* in this experiment.

In regard to the Stress Experience Inventory the analyses showed that anxiety and failure-stress items discriminated among the respective groups at the .001 level of significance. These results appear to validate the original anxiety classification and the particular set of instructions used to induce psychological stress, and this was the purpose of introducing the Stress Experience Inventory. The ego-involvement items did not prove to be sensitive to the original ego-involvement instructions. This may have been due in part to the items themselves or to complex situational conditions which occurred between the original administration of the ego-involvement condition and the Stress Experience Inventory. These results, however, serve to substantiate a basic assumption to the effect that the experimental conditions would serve as stressful experiences.

The results obtained for each of the objective measures of performance are clear and consistent. Flexibility as measured by the Woodworth-Wells test is significantly impaired by the stress conditions of the experiment. It is also worth noting that both error and block measures were independently sensitive to the stress conditions. Due to the conflicting elements of the Woodworth-Wells test, the sensitivity of error and blocking measures support the position that there is a loss of abstract ability under stressful conditions (**1, 10, 15**). The stresses employed in this study may be considered to have interfered with the voluntary ability to shift rapidly from one aspect of the situation to another.

Manifest anxiety does not appear to be related to performance until the situation becomes stressful. In this respect manifest anxiety may be interpreted as a "primer" or sensitizer to stress effects. This point has been developed in an earlier study in this series (**3**). These results do not appear to support the theoretical formulation of Taylor and Spence which was deduced from Hull's system to account for the apparent relationship between anxiety and ease of conditioning on one hand and decrement in complex performance tasks on the other (**18, 21**). In this formulation the higher the anxiety, the greater the drive strength and the consequent excitatory potential which gives rise to strength of response. In a single response situation, e.g., eyelid conditioning, the effect of anxiety is to facilitate conditioning (**19**). In learning and problem-solving, however, where a number of responses are required, the

increase in drive increases the excitatory potential of both relevant and irrelevant responses. The increase in excitatory potential of irrelevant responses decreases the probability of correct responses and results in performance decrement (**14, 16, 21**). According to this rationale the two anxiety groups in the present experiment should perhaps have been differentiated during the prestress trials. Because this differentiation was not found in the results, it would appear that the nature of the anxiety factor is mainly that of heightening the sensitivity of the more "anxious" *S*s to experimental stress.

Summary

This experiment was designed to determine the effects of two levels of anxiety, three levels of failure stress, and three levels of ego involvement on performance of a complex verbal-coding task. A total of 36 male and 18 female undergraduate students were selected as *S*s on the basis of high and low scores on a modified Taylor Anxiety Scale. The $3 \times 3 \times 2$ factorial design with three *S*s per cell permitted the assessment of individual differences as a separate source of variations among the results. Time, error, and frequency of blocking scores were used as measures of performance. Time scores were recorded for both pre- and poststress trials while errors and blocks were scored only on poststress trials. A Stress Experience Inventory was given to *S* after his testing session.

The results show that both errors and frequency of blocking are independently sensitive to the anxiety and stress conditions. On prestress trials the high- and low-anxiety groups were not found to differ significantly in terms of time taken to perform the test. On poststress trials the low-anxiety group showed a decrease in time while the high-anxiety group showed an increase in time. This was interpreted as a possible sensitizing function of anxiety to stress-producing situations.

References

1. Beier, E. G. The effect of induced anxiety on flexibility of intellectual functioning. *Psychol. Monogr.*, 1951, **65**, No. 9 (Whole No. 326).
2. Cattell, R. B., & Tiner, G. L. The varieties of structural rigidity. *J. Pers.*, 1949, **17**, 321–341.
3. Davidson, W. Z., Andrews, T. G., & Ross S. Effects of stress and anxiety on continuous high-speed color naming. *J. exp. Psychol.*, 1956, **52**, 13–17.
4. Deese, J., & Lazarus, R. S. The effects of psychological stress upon perceptual-motor performance. Lackland AFB, Tex.: Air Training Command Human Resources Res. Center, Res. Bull. 52–19, 1952.
5. Gaier, E. L. The relationship between selected personality variables and the thinking of students in discussion classes. *Sch. Rev.*, 1952, **60**, 404–411.
6. Holtzman, W. H., & Bitterman, M. E. Psychiatric screening of flying personnel. VI. Anxiety and reactions to stress. *USAF Sch. Aviat. Med. Rep.*, 1952, Proj. No. 21-37-002 (Rep. No. 6).
7. Hurlock, E. B. The value of praise and reproof as incentives for children. *Arch. Psychol., N. Y.*, 1924, **11**, No. 71.
8. Johnson, P. O., & Tsao, F. Factorial design and covariance in the study of individual educational development. *Psychometrika*, 1945, **10**, 133–162.
9. Katchmar, L. T. A review and analysis of the concept of stress and related variables. *Univer. of Maryland, Tech. Rep.*, 1954, Proj. No. DA 49-007-MD-222 (Rep. No. 16).
10. Krech, D., & Crutchfield, R. S. *Theory and problems of social psychology.* New York: McGraw-Hill, 1948.
11. Lantz, B. Some dynamic aspects of success and failure. *Psychol. Monogr.*, 1945, **59**, No. 1 (Whole No. 271).
12. Lazarus, R. S., Deese, J., & Osler, S. J. The effects of psychological stress upon performance. *Psychol. Bull.*, 1952, **49**, 293–317.
13. Lazarus, R. S., & Erickson, C. W. Effects of failure stress upon skilled performance. *J. exp. Psychol.*, 1952, **43**, 100–105.
14. Lucas, J. D. The interactive effects of anxiety, failure and intraserial duplication. *Amer. J. Psychol.*, 1952, **65**, 59–66.
15. Lybrand, W. A. The effects of certain psychological stress conditions on reasoning performance. Unpublished Doctor's dissertation, Univer. of Maryland, 1954.
16. Maltzman, I., Fox, J., & Morrisett, L. Some effects of manifest anxiety on mental set. *J. exp. Psychol.*, 1953, **46**, 50–54.
17. Miller, J. G., Bouthilet, L., & Eldredge, C. A bibliography for the development

of stress sensitive tests. PRB Tech. Res. Rep. No. 22, Oct. 1953.

18. SPENCE, K. W., & TAYLOR, J. A. Anxiety and strength of the unconditioned stimulus as determiners of the amount of eyelid conditioning. *J. exp. Psychol.*, 1951, **42,** 183–188.
19. TAYLOR, J. A. The relationship of anxiety to the conditioned eyelid response. *J. exp. Psychol.*, 1951, **41,** 81–92.
20. TAYLOR, J. A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48,** 285–290.
21. TAYLOR, J. A., & SPENCE, K. W. The relationship of anxiety level to performance in serial learning. *J. exp. Psychol.*, 1952, **44,** 61–64.
22. WESLEY, E. L. Perseverative behavior in a concept formation task as a function of manifest anxiety and rigidity. *J. abnorm. soc. Psychol.*, 1953, **48,** 129–134.
23. WOODWORTH, R. S., & WELLS, P. L. Association tests. *Psychol. Monogr.*, 1911, **13,** No. 5 (Whole No. 57).
24. VERVILLE, E. The effect of emotional and motivational sets on the perception of incomplete pictures. *J. genet. Psychol.*, 1946, **69,** 133–145.

(Received June 28, 1957)

Journal of Experimental Psychology
Vol. 55, No. 6, 1958

MEDIATED STIMULUS EQUIVALENCE AND DISTINCTIVENESS IN HUMAN CONDITIONING [1]

G. ROBERT GRICE AND JOHN D. DAVIS

University of Illinois

The role of cue-producing responses in the mediation of stimulus equivalence and distinctiveness has received considerable theoretical attention, but the experimental attack on the problem has been surprisingly small in relation to its presumed importance. Having its modern origins in Hull's concept of the "pure stimulus act" and his account of "secondary generalization" (**6**) and in Guthrie's theory of generalization (**4**), this idea has been elaborated in several contexts. Among the more important subsequent theoretical treatments of the problem are those of Cofer and Foley (**1**), Hilgard and Marquis (**5**), Dollard and Miller (**2, 11**), Osgood (**12**), and Goss (**3**). The essence of the view is that stimuli, to which some common response is made, tend to become functionally equivalent in evoking further responses. Conversely, stimuli to which different responses are made tend to become less equivalent or more distinctive. This acquired equivalence or distinctiveness is assumed to be mediated by supplementary cues produced by the common or differential responses.

Much of the work in this area has dealt with behavior of a rather high level of complexity and the cue-producing responses have been regarded more often as inferred entities than as variables under experimental control (e.g., **7, 8, 12**). The experiments reported here represent an attempt to investigate the problem of mediated stimulus equivalence and distinctiveness at the level of conditioning and with experimental manipulation of the cue-producing responses. Two experiments by Shipley (**13, 14**) and one by Lumsdaine (**10**) have indicated that the conditioned eyeblink can serve as a mediating response in the evocation of a finger reaction. However, the blink response is quite limited with respect to the amount of experimental control possible, and a greater variety of mediating reactions is required if mediated distinctiveness as well as equivalence is to be investigated. In the present studies, the "cue-producing" responses are instructed manual reactions, and the response indicating the degree of generalization is the conditioned eyeblink.

The basic design involves a procedure of differential conditioning. There are three clearly discriminable tones varying in frequency which may be called "High," "Medium," and "Low." The medium tone is the CS, always paired with the UCS, a puff of air. The high and low tones are negative stimuli never paired with the UCS. Now if, during conditioning, *S* is instructed to make one manual response to the CS and one of the negative tones and a different manual response to the second negative tone, a paradigm for both mediated equivalence and distinctiveness is provided. Equivalence should be mediated between the CS and the negative stimulus to which

[1] This research was supported by the University Research Board, University of Illinois.

the same response is made, and distinctiveness should be mediated between the two negative tones and between the CS and the tone to which the response is different. In other words, the generalization gradient should be asymmetrical.

Data from four experiments are reported. Experiments 1 and 3 involve the above paradigm and differ in that the mediating responses occur concurrently with the conditioning trial in Exp. 1, but not until the trial is over in Exp. 3. Much of the theoretical writing indicates that the mediating responses need not be overt. In the present situation, one way this can be investigated is to present the differential conditioning trials when *S* is set to make the mediating response but before he actually makes it. Experiment 2 was conducted under the conditions of Exp. 1 except that a single manual response was made to all stimuli, and presumably involved only the mediated equivalence effect. Experiment 4 is a simple differential conditioning experiment involving no manual responses.

Method

Apparatus

The *S* was seated in a straight chair in a sound-deadened, sound-proofed room approximately 5 ft. by 7 ft. and 7 ft. high. His hands and arms rested conveniently on a small table. Lever switches with which *S* made the manual responses were mounted 12 in. apart in the top of the table. One switch was operated by pulling it toward *S* with the index finger of the right hand. The other was operated by pushing it to the left with the thumb of the left hand. The *S* was seated facing the narrow end of the room with his head approximately 5 ft. from the wall and was centered between the two side walls. Illumination was supplied by a 40-w. bulb in the center of the ceiling. Incident illumination measured at the position of *S*'s eyes was approximately 2.5 ft.-candles.

Eyeblink responses were recorded by the electrical mechanical system described by Spence (**15**). The recording equipment was constructed by the Hunter Mfg. Company, Iowa City, Iowa. The auditory stimuli were produced by a Hewlett-Packard, Model 205 AG audio-oscillator and were presented through a Permo-flux loudspeaker, Model 8UP-8-1, which rested on the floor by the wall opposite *S*. The three tones were 240, 850, and 1900 cps. Intensity measurements were made with the "B" scale of a General Radio Sound Level Meter, Type 759-B. During measurement *S* was seated in the room and readings were taken on each side of the head next to the ear. The average intensity of each of the two extreme tones was 59.5 db and the average intensity of the middle tone was 60.5 db. Since the B scale compensates for the loudness function of the ear at this level, the three tones may be regarded as approximately equal in loudness. The UCS was a puff of air of 1 lb./sq. in. pressure in Exp. 1 and 2, and 2 lb./sq. in. in Exp. 3 and 4. It was controlled by a pressure-reducing valve and released by a solenoid valve. The duration of the tones was 500 msec. The UCS was presented simultaneously with the termination of the CS and was limited to 40 msec. duration. All time intervals were controlled by Hunter electronic timers. An intercommunication system made conversation between *E* and *S* possible.

Procedure

Experiment 1.—The instructions read to *S* were essentially those of a discrimination reaction time experiment. A warning buzzer sounded at a variable interval of 2, 3, or 4 sec. before the tone. At the buzzer, *S* was instructed to look at a 1-in. white spot on a black background on the wall just above eye level, and to blink once. He was instructed to snap the appropriate switch in response to the tones. Both speed and accuracy were emphasized. In order to achieve complete counterbalancing of the high and low negative tones and of the right and left manual responses, four groups were necessary. They were told to respond as follows: Group RRL, right switch for the low and middle tones and left switch for the high tone; Group RLL, right switch for the low tone and left for the medium and high tones; Group LLR, left switch for the low and medium tones and right for the high tone; Group LRR, left switch for the low tone and right for the medium and high tones. The usual neutral instructions concerning the eyeblink were included with the reaction time instructions. Before conditioning, each tone was presented once, *E* telling *S* before each one which tone it was to be. The *S* was asked if he wanted to hear the sequence again, and if so,

it was presented a second time. During Trials 1–40 there were 20 conditioning trials to the middle tone and 10 presentations of each of the negative tones. During Trials 41–100 there were 20 additional conditioning trials and 20 presentations of each negative tone. The different tones were presented in a prearranged irregular order. During the last 60 trials each successive block of 15 contained 5 presentations of each tone. The intertrial intervals were equally often—15, 20, and 25 sec. in an irregular order. During Trials 1–20 *S* was told each time he snapped the wrong switch. From Trial 21 on, he was told only if he made an error to the same tone twice in succession.

CR's were counted to all stimuli in the latency range from 150 to 500 msec. With the gain setting employed, any deflection of the pen of 1 mm. or more was counted as a CR if it was not a continuation of a response starting in the 0–150 msec. The *Ss* giving more than half of their CR's in the range 150–300 msec. were to be eliminated as voluntary responders (**16**). Actually, no *Ss* in the present experiment met this criterion.

The *Ss* were 24 volunteers from a course in general psychology. There were 11 men and 13 women. Four additional *Ss* who made fewer than 8 CR's were eliminated. One *S* who made 43 errors in the reaction-time task was eliminated. For the 24 included, the number of errors ranged from 0 to 13, with a mean of 5.5 and a median of 4.5

Experiment 2.—The task was described by the instructions as a simple reaction time experiment, half of the *Ss* making the response with the right hand and half with the left. There were 20 *Ss*, 10 men and 10 women. The voluntary responder criterion was not employed in this experiment, although approximately half of the *Ss* met the criterion. As may be seen in a comparison of the results of Exp. 1 and 2 with those of Exp. 3 and 4 below, the reaction time instructions have a marked facilitative effect on the number of eyeblinks. As was to be expected, an inspection of the reaction times of the manual responses indicated that they were shorter for the simple reaction task than for the discrimination conditions. It therefore appears probable that the occurrence of a larger number of blinks in the 150–300 msec. range did not represent voluntary responding, but was produced by the earlier facilitative effect of the manual responses. Parenthetically, it may be added that a separate examination of the data indicated no difference in the result between *Ss* who did and who did not meet the voluntary responder criterion.

Experiment 3.—A pilot study indicated that if this experiment were conducted under conditions identical to those of Exp. 1 except for a delay of the manual responses, there would be such a low level of conditioning that it would be impossible to study the phenomenon. This was presumably because of the absence of the facilitating effect of the reactions made concurrently with the conditioning trials. Two procedural changes were made to increase the level of conditioning. The intensity of the UCS was raised from 1 to 2 lbs./sq. in. and Trials 1–40 were converted entirely to reinforced presentations of the CS. The reaction time instructions were eliminated, and *S* was told to signal which tone had occurred after it had gone off. An additional reminder of the manual response instructions was given after Trial 40. The *Ss* were 19 men and 21 women. Twenty-one *Ss* who made fewer than 8 CR's and 2 who met the voluntary responder criterion were eliminated. The number of manual response errors ranged from 0 to 10 with a mean of 2.6 and a median of 2.0.

Experiment 4.—The procedure was the same as in Exp. 3, with the exception of the absence of the manual reactions and a corresponding deletion from the instructions. The *Ss* were 10 men and 10 women. Eleven additional *Ss* who made fewer than 8 CR's and 4 who met the voluntary responder criterion were eliminated.

Results

The data presented are the numbers of CR's to each of the three stimuli during Trials 41–100. The means for Exp. 1 and 3 are presented in Table 1 and those for Exp. 2 and 4 are in Table 2. The summary of an analysis of variance of the results of Exp. 1 is presented in Table 3. The main Generalization effect ap-

TABLE 1

Mean Number of CR's to Each Stimulus in Exp. 1 and 3

Stimulus	Exp. 1 (N = 24)		Exp. 3 (N = 40)	
	Mean	*SD*	Mean	*SD*
"Same Resp." Stim.	11.29	5.48	7.38	3.88
CS	9.63	5.13	6.65	3.96
"Diff. Resp." Stim.	8.67	5.55	6.43	4.25

TABLE 2

MEAN NUMBER OF CR'S TO EACH STIMULUS IN EXP. 2 AND 4

Stimulus	Exp. 2 (N = 20)		Exp. 4 (N = 20)	
	Mean	*SD*	Mean	*SD*
Low	15.9	5.30	5.45	3.28
Medium (CS)	15.0	5.02	7.65	4.37
High	15.0	6.02	6.10	3.90

pears as a within-Ss effect. Three categories appear under this effect—number of CR's to the CS, number to the negative stimulus with the same response, and number to the negative stimulus with the different response. Within the two negative stimulus categories, the specific tones and the specific manual responses are counterbalanced. The middle tone is, of course, always the CS. The use of the right or the left hand as the "same" response appears as a between-Ss effect, as does the use of the high or low tone as the stimulus to which "same" response is made. A further property of the present design is that an effect attributable to the high and low tones themselves would appear as a significant Generalization × Tone arrangement interaction. Similarly, a significant right- or left-hand effect would appear as a significant Generalization × Hand arrangement interaction (**9**, Type 3, p. 281). While there is little interest in the over-all value of F for the Generalization effect in the present experiment, it was significant at the .01 level. As predicted, the number of responses to the Same Response negative stimulus was greater than to the Different Response negative stimulus. A t of 3.28 based on the within-Ss error indicated that this difference is significant at the .01 level. Surprisingly, the number of responses to the Same Response stimulus was greater than to the CS itself. A t of 2.08 indicated that this difference is significant at the

TABLE 4

ANALYSIS OF VARIANCE FOR EXP. 2

Source	*df*	*MS*	*F*
Between *Ss*	19		
Hand (H)	1	72.6	.80
Error (b)	18	90.2	
Within *Ss*	40		
Tone (T)	2	5.4	2.16
H × T	2	2.4	.96
Error (w)	36	2.5	
Total	59		

TABLE 3

ANALYSIS OF VARIANCE FOR EXP 1

Source	*df*	*MS*	*F*
Between *Ss*	23		
Tone arrangement (T)	1	32.0	.38
Hand arrangement (H)	1	56.9	.68
T × H	1	4.5	.05
Error (b)	20	83.2	
Within *Ss*	48		
Generalization (G)	2	42.5	5.59
G × T	2	1.1	.14
G × H	2	6.6	.87
G × T × H	2	8.4	1.11
Error (w)	40	7.6	
Total	71		

TABLE 5

ANALYSIS OF VARIANCE FOR EXP. 3

Source	*df*	*MS*	*F*
Between *Ss*	39		
Tone arrangement (T)	1	50.7	1.33
Hand arrangement (H)	1	28.1	.74
T × H	1	61.6	1.62
Error (b)	36	38.0	
Within *Ss*	80		
Generalization (G)	2	9.9	1.71
G × T	2	13.4	2.31
G × H	2	2.0	.34
G × T × H	2	1.9	.33
Error (w)	72	5.8	
Total	119		

TABLE 6

ANALYSIS OF VARIANCE FOR EXP. 4

Source	*df*	*MS*	*F*
Between *Ss*	19		
Sex (S)	1	6.7	.19
Error (b)	18	34.6	
Within *Ss*	40		
Tone (T)	2	25.6	3.51
S × T	2	5.0	.68
Error (w)	36	7.3	
Total	59		

.05 level. The difference between the CS and the Different Response stimulus was not significant.

The summary of the analysis of Exp. 2 is presented in Table 4. In the absence of the differential responding variable, generalization now appears as the simple Tone effect within *Ss*. No value of *F* or *t* was significant.

The analysis of the data of Exp. 3 is presented in Table 5. No value of *F* was significant in this analysis. In view of the theoretical predictions and the results of Exp. 1, a *t*-test was applied to the difference between the two negative stimuli. The value of 1.8 is significant at the .05 level when viewed as a one-tailed test. The other differences are not significant.

The analysis of Exp. 4 is summarized in Table 6. The *F* for the Tone effect was significant at the .05 level. No other effect was significant. Two *t* tests of the one-tailed hypothesis that the number of responses to the negative stimuli were less than to the CS were significant at the .01 level for the low tone and at the .05 level for the high tone. The numbers of responses to the two negative stimuli did not differ significantly.

DISCUSSION

The significant difference in the predicted direction between the numbers of responses to the two negative stimuli in Exp. 1 indicates that a mediational effect was produced and lends support to the theoretical mechanisms under consideration. In two respects the results were not in accord with prediction. First was the significantly greater-than-100% generalization to the "equivalent" negative stimulus. Second was the absence of a significant difference between the CS and the "distinctive" negative stimulus. While the numbers to CR's to the negative stimuli bore the predicted relation to each other, *both* were greater than predicted with respect to the CS. This suggests the predicted operation of the mediation effect complicated by an additional factor acting to increase the level of response to the negative stimuli. One possibility is the facilitative effect of the manual responses. The negative stimuli were the two extremes of the frequency range, which implies that they were the most recognizable or identifiable. The result may have been more vigorous manual reactions, and hence, a greater facilitive effect for these tones than for the CS.

The flat generalization gradient of Exp. 2 is in accord with the fact that only mediated equivalence was involved. The failure to obtain generalization significantly greater than 100%, as in the corresponding portion of Exp. 1, tends to support the differential facilitation interpretation, since there was no requirement to discriminate or respond differentially. If the mediated equivalence effect alone produces a flat gradient, the difference between the CS and the Same Response stimulus of Exp. 1 may be taken as an estimate of the amount by which response to *both* negative stimuli were differentially facilitated. The result of subtracting this amount from the numbers of responses to the negative stimuli may be seen in the middle curve of Fig. 1. Tentatively, this may be regarded as a mediated generalization gradient from which the differential facilitation effect has been eliminated—the left wing illustrating mediated equivalence and the right wing mediated distinctiveness. The

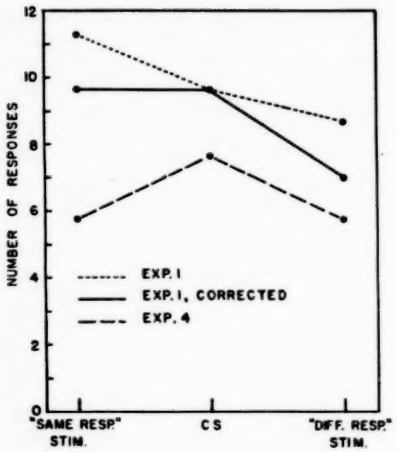

FIG. 1. Uncorrected and corrected generalization gradients for Exp. 1, and generalization gradients for Exp. 4 with the high and low tones averaged.

form is in general agreement with the initial prediction.

The results of Exp. 3 fail to clearly demonstrate a mediational effect from a set to respond. However, the similarity of the trend to that of Exp. 1 (cf. Table 2), and the value of the one-tailed t between the negative stimuli are promising and suggest that the effect may ultimately be found to be genuine.

The data from Exp. 4 indicate that significant differential conditioning is obtained in this situation in the absence of experimentally manipulated mediating or facilitating responses. Particularly in view of the greater number of conditioning trials and the stronger UCS, the smaller number of CR's is evidence for the facilitation effect in the first two experiments.

SUMMARY

In an investigation of the mediating effects of cue-producing responses, four experiments were conducted employing differential eyeblink conditioning in which the middle of three tones varying in frequency was the conditioned stimulus. The *S*s also made instructed, manual responses to the tones according to a paradigm designed to produce mediated stimulus equivalence and distinctiveness.

There were more conditioned eyeblinks to negative stimuli to which *S*s made the same manual response as to the CS than to those to which they made a different response. The results are interpreted as being in general agreement with a theory of response mediated stimulus equivalence and distinctiveness.

The mediating effect was most pronounced if the mediating responses occurred overtly during the conditioning trial. The data also suggest, but not conclusively, that a set to respond may have a mediating effect.

The results were somewhat complicated by what appeared to be a direct facilitative effect of the manual responses on the eyeblink.

REFERENCES

1. COFER, C. N., & FOLEY, J. P. Mediated generalization and the interpretation of verbal behavior: I. Prologomena. *Psychol. Rev.*, 1942, **49**, 513–540.
2. DOLLARD, J., & MILLER, N. E. *Personality and psychotherapy.* New York: McGraw-Hill, 1950.
3. GOSS, A. E. A stimulus-response analysis of the interaction of cue-producing and instrumental responses. *Psychol. Rev.*, 1955, **62**, 20–31.
4. GUTHRIE, E. R. *The psychology of learning.* New York: Harper, 1935.
5. HILGARD, E. R., & MARQUIS, D. G. *Conditioning and learning.* New York: D. Appleton-Century, 1940.
6. HULL, C. L. The problem of stimulus equivalence in behavior theory. *Psychol. Rev.*, 1939, **46**, 9–30.
7. LAWRENCE, D. H. Acquired distinctiveness of cues: I. Transfer between discriminations on the basis of familiarity with the stimulus. *J. exp. Psychol.*, 1949, **39**, 770–784.
8. LAWRENCE, D. H. Acquired distinctiveness of cues: II. Selective association in a constant stimulus situation. *J. exp. Psychol.*, 1950, **40**, 175–188.
9. LINDQUIST, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953.
10. LUMSDAINE, A. A. Conditioned eyelid responses as mediating generalized finger reactions. *Psychol. Bull.*, 1939, **36**, 650. (Abstract)

11. Miller, N. E., & Dollard, J. *Social learning and imitation.* New Haven: Yale Univer. Press, 1941.
12. Osgood, C. E. *Method and theory in experimental psychology.* New York: Oxford Univer. Press, 1953.
13. Shipley, W. C. An apparent transfer of conditioning. *J. gen. Psychol.*, 1933, **8,** 382–391.
14. Shipley, W. C. Indirect conditioning. *J. gen. Psychol.*, 1935, **12,** 337–357.
15. Spence, K. W. Learning and performance in eyelid conditioning as a function of the intensity of the UCS. *J. exp. Psychol.*, 1953, **45,** 57–63.
16. Spence, K. W., & Taylor, J. A. Anxiety and strength of the UCS as determiners of the amount of eyelid conditioning. *J. exp. Psychol.*, 1951, **42,** 183–188.

(Received July 5, 1957)

Journal of Experimental Psychology
Vol. 55, No. 6, 1958

SERIAL VERBAL LEARNING UNDER TWO CONDITIONS OF HUNGER MOTIVATION[1]

ROBERT G. LERNER, IRWIN SINGER, AND HARRY C. TRIANDIS

Cornell University

The influence of affective processes in perception has been subject to intense experimentation in the last ten years. Jenkin (**5**) in a recent review of more than one hundred papers indicates the wealth of experimentation in this field and suggests that a satisfactory theoretical explanation of the phenomena is still lacking. One attempt to provide a theoretical explanation has been made by Bruner (**3**) and by Postman (**6**). The Bruner-Postman theory was further elaborated by Allport (**2,** pp. 407–436) who recast it in terms of his theory of event structure (**1**) and called it a "structural theory of set dynamics." The present paper presents data which are consistent with the Allport theory.

The influence of hunger on some aspects of perception seems well established. Jenkin writes: "Of the six reports upon hunger in relation to a broadly defined cognitive area, only one was essentially negative, and this was the one which used stimuli largely irrelevant to the need-state studied. It seems from the empirical findings that when appropriate stimulus objects are used, the presence of the hunger need facilitates categorization in a manner consistent with that need" (**5,** p. 107). In the present study hunger was selected as the motivational variable because of its relatively well established relationship to perceptual phenomena. Allport's analysis of the phenomena of the influence of hunger on perception includes the concept of set—hungry *S*s are "more set to be aware of food items" (**2,** p. 415)—and describes the hungry *S* as being in a situation which affords biasing conditions for the appearance of a certain structure because hunger contractions (ongoings in the stomach) are already supporting a part of the (potential) format of the structure.

One of the more satisfactory theories which account for the phenomena of verbal learning is the one proposed by Gibson (**4**). She suggests that verbal learning involves the establishment of discrimination among the various items. She also argues that *meaning*—the "characteristic of a verbal item which serves to differentiate it from other items"—plays a vital role in verbal learning.

Simultaneous consideration of Allport's notion of greater awareness of food items for hungry *S*s than for satiated ones, and Gibson's views on differentiation leads to the prediction that if a list of very common words, including food words, is presented to hungry *S*s, the food words would be subject to greater differentiation than the nonfood words. In the case of the control (satiated) group, there will be no such differentiation. Thus, in the case of the hungry group, the food words will be more identifiable; or in Gibson's terms, will generalize less. In the special case of serial learning, each word is a stimulus for the word that follows it. A better differentiated word should be able to function as a better cue for the *next*

[1] The authors wish to thank Eleanor J. Gibson for her critical reading of the manuscript.

word on the list. These considerations lead to the prediction that in a serial-learning situation the words which *follow* the food words will be learned first. The results of our experiment confirm this prediction.

Method

Apparatus.—A memory drum was used. Each word was viewed for 1 sec. and was followed by a blank space which remained in view for 1 sec. The anticipation method was used.

Subjects.—An experimental group consisting of 20 "hungry" *S*s and a control group consisting of 20 "satiated" *S*s were used. For a period of six months prior to testing, all *S*s ate dinner at 6:10 P.M. daily and were finished by 6:45. Lunch was always over at 1:20 P.M. The experimental group was administered the test between 5 and 6 P.M. Each *S* took about 20 min. The control group took the same test between 8 and 9 P.M.

Procedure and list of words.—Ten trials, using the anticipation method, were used per *S*. The *S* was informed that there would be 10 trials during the test period. He was instructed to try to learn the list using any mnemonic device he desired. The list consisted of very common one-syllable words. There were 19 nonfood and 8 food words. They were the following: Pen, Nail, Shirt, *Glass*, Red, Train, *Bread*, Cat, Bush, *Steak*, Paint, Door, *Cake*, Ink, Green, *Spoon*, Dog, Bus, *Milk*, Drum, Blue, *Fork*, Desk, Car, *Tea*, Dark, Tree.

Results

Tables 1 and 2 show the results obtained from the two groups. The hungry *S*s learned the words which follow the food words more readily

TABLE 1

Correct Responses to Words When Food Words Were Stimuli

Exp. Group (N = 20)		Control Group (N = 20)	
F. W. as R	F. W. as S	F. W. as R	F. W. as S
260	410	232	210

Note:—In the hungry group, eight of the eight food words as stimuli lead to more correct responses than in the satiated group. In the hungry group, five out of eight food words were called more correctly than in the satiated group.

TABLE 2

Correct Responses per *S* per Word

Measure	Experimental		Control		P_{diff}
	Mean	*SD*	Mean	*SD*	
Food words as R's	1.69	2.20	1.45	2.11	>.05
Food words as S's	2.56	2.88	1.31	2.18	<.001
Other words as R's	3.74	3.05	3.52	2.87	>.05
All Words	2.76	3.75	2.26	3.43	<.01

than any of the other words. The mean numbers of correct responses, when the food words were the stimuli, were 20.5 per *S* for hungry *S*s and 10.5 per *S* for satiated *S*s. The corresponding numbers of correct responses per *S* per word were 2.56 and 1.31. The difference is significant at the .001 level, with a nonparametric test (randomization test, **8**). For the nonfood words, the mean numbers of correct responses per *S* were 54.1 and 50.35, respectively; the difference between these two means is not statistically significant.

An analysis of the errors of response of the two groups indicated that about 75% of the errors were food responses, in the case of the hungry group; in the case of the satiated group only about 30% of the errors were food responses, which is a figure that one would expect by chance. In general, when the food word was the cue, *S*s in the hungry group responded with a nonfood word, which is further evidence that they differentiated the two kinds of words. One may consider this as evidence that *S*s "realized" that at no time did a food word follow a food word. This suggests that *S*s were able to categorize the words into "food" and "nonfood" words. The control group did not use this categorization.

Discussion

It seems clear that selective sensitization (**7**) or the operation of a hunger-structure (**2**) can be shown to be a factor in the serial learning of hungry *S*s. In the present experiment we did not depend on reduced cues, such as is the case in tachistoscopic experiments, or on ambiguous figures, such as those used to demonstrate the phenomenon of autism in perception. Most of the methodological criticisms of the experiments on the influence of affective processes in perception (**2, 5**) do not apply to the present experiment.

The serial-position effect can be seen to operate in the within-groups data. Since the nonfood words were situated at the end of the series the number of correct responses to them is much higher than in the case of the food words, which were located more towards the middle of the series. Thus, the serial position effect operates to decrease the effects of selective sensitization, in the case of the present experiment, so that the findings are not likely to be due to an artifact.

Summary

Twenty hungry and 20 satiated *S*s were used in a serial verbal-learning experiment, with a list of 8 food words and 19 nonfood words. The hungry *S*s learned the words which follow the food words more readily than any of the other words. The phenomenon is derived from Allport's structural theory of set dynamics, and Gibson's theory of generalization and differentiation in verbal learning.

References

1. Allport, F. H. The structuring of events: Outline of a general theory with applications to psychology. *Psychol. Rev.*, 1954, **61**, 281–303.
2. Allport, F. H. *Theories of perception and the concept of structure.* New York: Wiley, 1955.
3. Bruner, J. S. Personality dynamics and the process of perceiving. In R. R. Blake and G. V. Ramsey (Eds.), *Perception: An approach to personality.* New York: Ronald Press, 1951. Ch. 5.
4. Gibson, E. J. A systematic application of the concepts of generalization and differentiation to verbal learning. *Psychol. Rev.*, 1940, **47**, 196–229.
5. Jenkin, N. Affective processes in perception. *Psychol. Bull.*, 1957, **54**, 100–127.
6. Postman, L. Towards a general theory of cognition. In J. H. Rohrer and M. Sherif (Eds.), *Social psychology at the crossroads.* New York: Harper, 1951. Ch. 10.
7. Postman, L., Bruner, J. S., & McGinnies, E. Personal values as selective factors in perception. *J. abnorm. soc. Psychol.*, 1948, **43**, 142–154.
8. Siegel, S. *Nonparametric statistics.* New York: McGraw-Hill, 1956.

(Received July 11, 1957)

Journal of Experimental Psychology
Vol. 55, No. 6, 1958

ACHIEVEMENT MOTIVATION AND GOAL-SETTING BEHAVIOR ON A LEARNING TASK

DONALD H. KAUSLER AND E. PHILIP TRAPP

University of Arkansas

The McClelland group (**7**) has reported studies on nonlearning tasks which indicate that level of aspiration is determined, in part, by level of achievement motivation (*n* Achievement). That is, *S*s with a high *n* Achievement are likely to set higher levels of aspiration than are *S*s with low *n* Achievement. McClelland suggests that this relation holds true as long as the reality determinants of level of aspiration are uncertain or in conflict. The reality determinant in the level of aspiration situation is essentially *S*'s knowledge of his past performance on the task. McClelland further suggests that when the reality determinants are consistent and unambiguous, level of aspiration is predominantly controlled by these determinants, with *n* Achievement dropping out as a critical factor.

The present study tests the McClelland hypothesis with respect to goal-setting behavior in a learning situation. Subjects with high *n* Achievement would be predicted to set higher levels of aspiration on learning tasks than *S*s with low *n* Achievement. The difference between high and low achievement groups, however, should diminish during a sequence of learning trials if *S*s are aware of their progress. As knowledge of results accumulate, *S*s should begin to formulate a more realistic evaluation of their ability on the task and their rate of progress on the task. It is hypothesized that this modified frame of reference, once established, would operate to decrease the differential effects of *n* Achievement on goal-setting behavior.

Several recent studies (**1, 6, 8**) have indicated that performance level may also be differentially affected by achievement motivation. In order to control the possible effects of performance differences on the magnitude of aspiration scores, goal-discrepancy scores (D-scores) served as the measure of goal-setting behavior in the present study. The specific hypotheses investigated were: (*a*) *S*s with a high level of achievement motivation will have significantly higher D-scores than will *S*s with a low level of achievement motivation; and (*b*) the difference between the two groups in D-scores will diminish with practice on the learning task when *S*s are informed of their progress.

METHOD

Subjects.—The *S*s were 83 students, approximately the same number of each sex, in abnormal psychology at the University of Arkansas. The study was conducted during regular class periods.

Instruments.—The test developed by French (**2, 3, 4, 5**) to measure complex motivations was used as the achievement-motivation measure. The test was scored for achievement motivation by the system suggested by French. The authors found close agreement between test scores for two scorers working independently.

An extended group form of the Wechsler-Bellevue Digit Symbol Test served as the learning task. The test was prepared in booklet form for group administration. The first page contained, in addition to space for identification data and a practice trial, spaces for *S*s to record both their level of aspiration and their performance score (number of correctly filled boxes) for each trial. Each of the remaining five pages of the booklet replicated the Wechsler test extended in length to 75 unfilled boxes.

Procedure.—Early in the semester *S*s were administered the French test which was disguised as a "Test of Insight." They were divided into two groups, a high *n* Achievement

and a low *n* Achievement group, on the basis of their test scores. The *Ss* scoring above the median for the total number of *Ss* were assigned to the high *n* Achievement group ($N = 41$); the *Ss* scoring below the median were assigned to the low *n* Achievement group ($N = 42$). This procedure of assigning *Ss* follows that used by French in her studies involving the test **(2, 5)**.

The digit-symbol test was administered a month after the French test. Apparently none of the *Ss* suspected any connection between the use of the two tests. The *Ss* were informed that they were taking one of the subtests of the Wechsler test in an experimental group form to see if it could predict grades in college courses as well as the individual form does. The *Ss* were also informed that *E* was interested in seeing how well they could estimate their performance from trial to trial on the test.

Following the reading of Wechsler's standardized instructions for the test and the working of the sample items, *Ss* began the first of five trials, each with a 1-min. time limit. After each trial, *Ss* counted the number of correctly filled boxes and recorded this number in the appropriate space on the first page of the booklet. They then estimated their performance on the next trial. No performance estimate was made following Trial 5. The performance estimates served as a measure of level of aspiration. The time between trials was 2 min.

Results

For each *S* four D-scores were computed (level of aspiration for Trial 2 minus performance score for Trial 1, etc.). An analysis of variance of the data was made to determine both the over-all effect of *n* Achievement on D-scores and the difference in trend over the four trials for the two motivational conditions. The results of this analysis are given in Table 1.

TABLE 1

Analysis of Variance of D-Scores

Source	*df*	*MS*	*F*
Between high and low *n* Achievement	1	150.7	3.75
Between *Ss* in same group	81	40.2	
Between trials	3	85.0	
Trials × *n* Ach.	3	25.8	4.16*
Pooled *Ss* × Trials	243	6.2	
Total	331		

* $P < .01$.

TABLE 2

D-Score by Trial for High and Low *n* Achievement Groups

Trial	High *n* Ach.		Low *n* Ach.		*t*
	Mean	*SD*	Mean	*SD*	
1–2	5.59	4.43	2.98	4.51	2.64*
2–3	3.07	3.37	1.76	3.59	1.70
3–4	2.83	3.40	1.76	3.86	1.32
4–5	1.93	3.12	1.76	3.16	.25

* $P < .01$ (two-tailed test).

An average D-score for all four trials was determined for each *S*. For the high *n* Achievement group, the mean and *SD* of these average D-scores were 3.37 and 3.00. For the low *n* Achievement group, the mean and *SD* were 2.06 and 3.41. Although the difference is in the hypothesized direction, the variance attributable to *n* Achievement is not significant ($F = 3.75$, $df = 1$ and 81).

The variance attributable to the interaction between *n* Achievement and Trials is significant ($P < .01$, $F = 4.16$, $df = 3$ and 243). To analyze further the difference in trend between the two groups, the differences in mean D-scores between the two groups on each trial were tested for significance by the *t* test. The results of this analysis are given in Table 2. On the first trial the mean D-score for the high *n* Achievement group is significantly greater than that for the low *n* Achievement group ($t = 2.64$, $P < .01$). On the last three trials the difference between

group means is not significant. The results thus support the related hypothesis.

Discussion

The most pertinent finding was the demonstrated inverse relationship between *n* Achievement and the reality determiners on a level of aspiration task involving the learning process. With reality factors operating at a minimum, *S*'s expressed expectancy of performance, as measured by the D-score, was significantly influenced by his level of *n* Achievement. This influence, however, disappeared as *S* received knowledge of results. It was found that after one trial group differences were reduced to chance level. The results thus indicate two major findings with respect to goal-setting behavior. First, they demonstrate the significant role that *n* Achievement does play as a determiner of goal level in a level of aspiration situation. Secondly, they emphasize that *n* Achievement is suppressed as a goal determiner by the presence of reality checks during performance. Research on level of aspiration should take cognizance of *n* Achievement as an important variable in goal setting.

The study further suggests the value of more systematic research into the relationship between motivation and goal-setting behavior. *n* Achievement, being but one facet of human motivation, provides a clue as to how other motivational forces may interact with reality factors in the setting of specific goals. Other independent variables that have been described in the literature as affecting level of aspiration, such as social frames of reference, may actually be operating as instigators of *n* Achievement. That is, by manipulating these variables, *Es* may, in effect, be manipulating a motivational variable. A repeat of these studies involving the more clearly defined concept of *n* Achievement might lead to a reduction of the number of presumed independent variables related to goal-setting behavior.

Methodologically, the study supports the usefulness of the French achievement test for differentiating *Ss* on the *n* Achievement variable. Since the administration time is brief and its scoring quick and objective, it may be recommended for discriminating between groups on the *n* Achievement variable.

Summary

Two hypotheses derived from McClelland's studies on achievement motivation and level of aspiration were tested in a learning situation: (*a*) *Ss* with a high level of achievement motivation will have significantly higher *D*-scores than *Ss* with a low level of achievement motivation; (*b*) the difference between high and low *n* Achievement groups in D-scores will diminish with practice if *Ss* are informed of their progress.

Eighty-three *Ss* were given the *n* Achievement test developed by French, and the median was the cutting score for the high and low *n* Achievement groups. The learning task was an extended form of the Wechsler digit-symbol test and was given a month after the French test. The *Ss* were asked before each of five trials to estimate their performance score, and they counted and recorded their actual score after each trial.

The mean of average D-scores by *S* for all four trials was greater for the high *n* Achievement group than the mean for the low *n* Achievement group, but the over-all variance attributable to *n* Achievement was not significant. The two groups displayed a statistically significant difference in trend (Trials × *n* Achievement interaction) over the four trials. When a trial-by-trial analysis was made, the high *n* Achievement group had greater mean D-scores on each trial, but only on the first trial was the difference between group means statistically significant. Therefore, both hypotheses were confirmed.

References

1. Atkinson, J. W., & Reitman, W. R. Performance as a function of motive strength and expectancy of goal-attainment. *J. abnorm. soc. Psychol.*, 1956, **53**, 361–366.
2. French, E. G. Some characteristics of achievement motivation. *J. exp. Psychol.*, 1955, **50**, 232–236.
3. French, E. G. Development of a measure of complex motivation. *USAF Personn. Train. Res. Cent. Res. Rep. No. TN-56-48*, 1956.
4. French, E. G. Motivation as a variable in work-partner selection. *J. abnorm. soc. Psychol.*, 1956, **53**, 96–99.

5. French, E. G., & Chadwick, I. Some characteristics of affiliation of motivation. *J. abnorm. soc. Psychol.*, 1956, **52,** 296–300.
6. Lowell, E. L. The effect of need for achievement on learning and performance. *J. Psychol.*, 1952, **33,** 31–40.
7. McClelland, D. C., Atkinson, J. W., Clark, R. A., & Lowell, E. L. *The achievement motive.* New York: Appleton-Century-Crofts, 1953.
8. Wendt, H. W. Motivation, effort, and performance. In D. C. McClelland (Ed.), *Studies in motivation.* New York: Appleton-Century-Crofts, 1955.

(Received July 15, 1957)

Journal of Experimental Psychology
Vol. 55, No. 6, 1958

EFFECT OF KNOWLEDGE OF CONDITIONING UPON EYELID CONDITIONING

WALLACE R. McALLISTER AND DOROTHY E. McALLISTER[1]

Syracuse University

One of the procedural questions which *E*s must answer in conducting an experiment on eyelid conditioning is whether or not to inform *S* concerning the nature of the experiment. That knowledge of conditioning might have some effect upon performance is indicated by the fact that specific instructions have been found to affect conditioning markedly (e.g., **1; 2; 3,** p. 261 f.). In interpreting such studies, it has been suggested (**2, 3**) that with neutral instructions *S*s in eyelid-conditioning experiments tend to adopt a negative set; i.e., they tend to inhibit the making of anticipatory conditioned responses (CR's). Such a negative set might be adopted by *S* (*a*) to avoid appearing foolish by blinking to the conditioned stimulus (CS) before the unconditioned stimulus (UCS) is delivered, or (*b*) to comply to incorrectly interpreted instructions (possibly resulting from being told to fixate the locus of the CS at the ready signal). If such is the case, it seems reasonable that, if *S*s knew the principles of conditioning and realized that anticipatory CR's were natural and likely to occur, the inhibitory set might not be adopted. It was the purpose of this investigation to determine whether knowledge of the nature of eyelid conditioning would enhance conditioning performance.

METHOD

Subjects and design.—Sixty-four *S*s, 32 men and 32 women, from an introductory psychology course at Syracuse University participated in the experiment in partial fulfillment of a requirement of the course. Half of the men and of the women, randomly selected, were given instructions which supplied them with full knowledge of the nature of the conditioning process; the other half were not given this knowledge. Thus, there were four groups: men and women knowledge (K) groups and men and women no-knowledge (NK) groups. In the NK condition, an additional seven *S*s—four men and three women—were eliminated because it was found, during questioning at the end of the experiment, that they had pre-experimental acquaintance with conditioning and were able to so label the experiment. These *S*s reported having studied about conditioning techniques in other science courses, in some cases in high school.

Conditioning procedure.—Each *S* was seated in a moderately illuminated (3.2 ft.-candles), sound-treated room about 30 in. in front of a wooden shield painted flat black. Cut into the shield at about eye level was a circular aperture (1¼ in. in diameter) covered with opal glass and a Wratten filter (No. 47B) which allowed passage of a bluish light used for the CS. The *E* administered the conditioning trials from an adjacent sound-treated room.

The CS was a light of approximately .6 log apparent ft.-candles of 900 msec. duration. The UCS was a puff of air of approximately .5 lb./sq. in. generated by an air system consisting of an air compressor, a storage tank, a pressure-reducing valve, and a solenoid-operated air valve. The puff was delivered to *S*'s right cornea through a nozzle 1 mm. in diameter. The duration of the puff was 100 msec. and was overlapped by the CS. Approximately 40 msec. were required for the puff to reach *S*'s eye. Thus the CS-UCS interval was 840 msec. The onsets and durations of the CS and UCS were controlled by Hunter interval timers.

Briefly, the method of recording the eyelid response involved (*a*) a small plastic lever fastened to the eyelid with adhesive tape, (*b*) a light-weight microtorque potentiometer, the variable (rotating) arm of which was connected to the plastic lever by a thread, (*c*) a Hunter eyeblink amplifier, (*d*) an Edin 8002 penmotor, and (*e*) a Hunter paper puller.

Prior to the conditioning trials, three presen-

[1] The authors wish to express their appreciation to Phyllis Alberts and Virginia Verhey for aid in running *S*s.

tations of the CS alone were given to insure that *S* was following instructions and to determine whether *S*'s eyelid response was already conditioned to the stimulus. If *S* gave three CR's, he was eliminated. One *S* was discarded for this reason. Then, three presentations of the UCS alone were given to acquaint *S* with the nature of the puff. These preliminary stimuli were presented 10 sec. apart. The conditioning trials were begun about 1 min. later.

At random intervals of 3, 4, or 5 sec. before the CS was presented, the signal, "Ready," was spoken over an intercommunication system between the *S*'s and *E*'s room. At this signal *S* was instructed to fixate the aperture and to blink once. The trial was then given. Seventy trials were presented in this manner, the intertrial interval averaging 30 sec., varying in a fixed, irregular order of 25, 30, or 35 sec.

The performance measure used was the percentage, in blocks of 10 trials, of anticipatory CR's defined as an eyeblink 1 mm. or more in amplitude, occurring between 250 and 840 msec. after the onset of the CS. A criterion for identifying voluntary conditioners was established beforehand in accordance with generally accepted procedures (**5**). All CR's occurring between 250 and 460 msec. were considered to be voluntary responses. All *S*s with more than 50% of their CR's falling in this interval were to be classified as voluntary conditioners. No *S* met this criterion.

Instructions.—Immediately after being seated and having the equipment adjusted, instructions were read to *S*. The *S*s in the K condition were informed of the nature of conditioning and what might be expected during the course of the experiment. The relevant part of their instructions follows.

"This is an eyelid-conditioning experiment. Before we begin, I want to explain briefly what the eyelid-conditioning technique is. First you will see a light in this circle [point] for a brief period and then you will feel a puff of air in the corner of your right eye. The air will always cause your eye to blink, so the sequence is light, air, blink. After this is done a number of times, some people, but not all people, come to blink automatically after the light comes on but before the air hits the eye. They do not blink intentionally; in fact, they often aren't even aware that they do blink before the air hits their eye. When a person does come to blink automatically to the light, we say that conditioning has taken place. The blink is now made to the light whereas before conditioning took place, the blink only occurred to the puff of air. The sequence after conditioning is light, blink, air. Do you understand what I've said? To be sure you do understand what conditioning is, could you express briefly in your own words what I've said?"

At this point *S* was allowed to explain conditioning in his own words. The instructions were not continued until *S* had expressed (*a*) the sequence of events before conditioning, (*b*) the sequence after conditioning, and (*c*) the automaticity of the response. The instructions then continued as follows:

"I've told you all of this so that you can better understand that in the study of conditioning it is very important that the subject does not intentionally blink his eye or intentionally prevent his eye from blinking. Therefore, as a subject you can help us most by trying to forget all about your eye and its movements. Think about other things—work you have to do—anything at all."

The introductory portion of the instructions to the *S*s of the NK condition follows. "In this experiment we are concerned with the reaction of your eye to different kinds of stimulation, particularly how it will respond to a light when repeated at frequent intervals. In addition to the light, a puff of air will be delivered to the corner of your right eye which will make it blink. You can help us most by trying to forget all about your eye and its movements. Think about other things—work you have to do—anything at all."

The remainder of the instructions was identical for both the K and the NK conditions and gave the necessary information concerning procedural matters. Included in this portion of the instructions was the following statement. "Don't try to control the reactions of your eye to either the light or the puff of air. I know you can do that, but that isn't the purpose of the experiment. Just try to keep relaxed at all times and let your eye do whatever it wishes." This portion of the instructions was introduced in accordance with established procedures in an attempt to induce a neutral set in *S*. The present hypothesis was, of course, that under so-called neutral instructions *S*s adopt a negative set and that knowledge-of-conditioning instructions would eliminate it. It is possible that explicit instructions about conditioning would elicit a facilitatory set in the K group and not merely eliminate the negative one. However, the fact that there were no voluntary conditioners might be considered as evidence against such a view.

At the termination of the experiment, *S*s were questioned with respect to their understanding of the purpose of the experiment and were cautioned not to discuss the experiment with other members of the class.

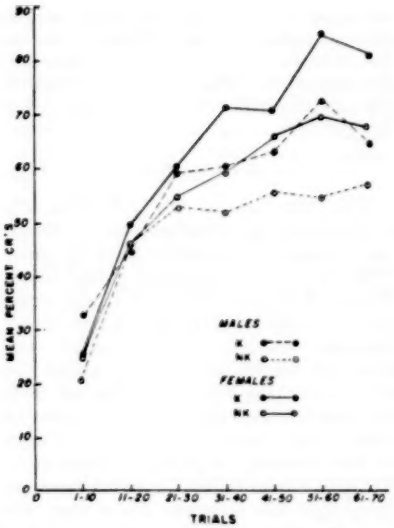

FIG. 1. Mean percentage of CR's as a function of trials for men and women under knowledge (K) and no-knowledge (NK) conditions.

RESULTS

In Fig. 1 is presented the conditioning performance for each of the four groups plotted as a function of blocks of 10 trials. It can be noted that the K group conditioned better than the NK group for each sex. It is also apparent that the women conditioned better than the men within each instruction condition. This latter result is consistent with previous findings in this laboratory.

The performance at the end of training was evaluated by means of a factorial analysis of variance[2] utilizing as the response measure the mean percentage of CR's over the last 20 trials. A summary table of the analysis and the means of the main factors are given in Table 1. As may be seen, neither the instructions factor nor the sex factor reached the level of significance generally used as the basis for rejecting the null hypothesis. However, each was significant at about the 8% level.

The lower level of conditioning manifested by the NK *Ss* was apparently brought about by the presence in this group of a greater number of *Ss* who failed to condition well. Six *Ss* (three of each sex) from the NK condition gave 20% or less CR's over the last 20 trials as compared to only two (both men) in the K condition. Further support is provided by a comparison of three sets of relatively homogeneous *Ss* selected from the K and NK conditions. The *Ss* were chosen by a procedure similar to one suggested by Spence (**4,** p. 68). The mean percentage of CR's over the 70 trials was computed for each *S*, and two frequency distributions were obtained, one for the K and one for the NK condition (sexes combined). The performance of the top, middle, and bottom 25% of the *Ss* in each distribution is plotted in Fig. 2 as a function of blocks of 10 trials.

TABLE 1

ANALYSIS OF VARIANCE OF PERCENTAGES OF CR's IN LAST 20 TRIALS

Source	Mean %	*df*	*MS*	*F*
Instructions (I)		1	2943.07	3.26*
Knowledge (K)	76.3			
No knowledge (NK)	62.8			
Sex (S)		1	2862.25	3.17*
Men	62.8			
Women	76.2			
I × S		1	12.25	<1
Within cells		60	903.74	
Total		63		

* $.05 < P < .10$.

[2] The assumption of homogeneity of variance was found to be tenable with Bartlett's test. The assumptions of randomness and normality were assumed to be met on the basis of the sampling procedure and observation of the obtained data, respectively.

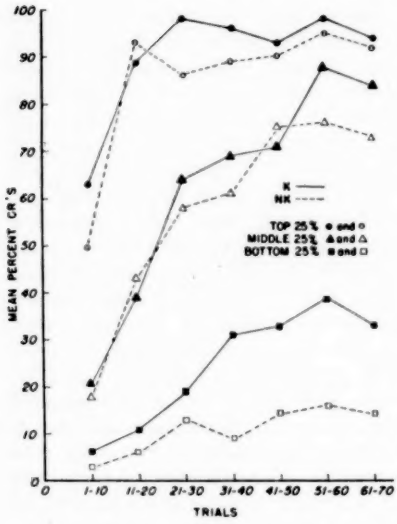

FIG. 2. Mean percentage of CR's as a function of trials for the top, middle, and bottom 25% of conditioners in the knowledge (K) and no-knowledge (NK) groups.

Clearly the difference between the curves for the K and NK conditions is greatest for the bottom 25% of the *Ss*. However, it will also be noted that, considering the three sets of curves, the NK mean exceeds the comparable K mean at only three points. Thus, while *Ss* at the three conditioning levels contributed to the obtained difference, the low conditioners showed the greatest effect.

A question might be asked regarding the effect of eliminating the seven *Ss* from the NK condition for having previous knowledge of conditioning in view of the fact that similar *Ss* from the K group were not discarded. This was the case because it was difficult to determine by the questioning after the experiment just how much actual pre-experimental knowledge and how much knowledge gained from the instructions entered into the answers of the K *Ss*. In any event, there would have been little difference in the results if the seven NK *Ss* had not been discarded. Adding their data to that of the NK condition (sexes combined) resulted in only slight changes in the performance curve. The mean change for the seven blocks of trials was only 1.2%. The mean difference over the last 20 trials for the main effect of instructions was reduced by 1.6%. Thus, it is not likely that the elimination of these *Ss* was of any consequence.

DISCUSSION

The direction of the difference obtained with the knowledge-of-conditioning variable is consistent with the view that explicit instructions about conditioning lead to a reduction in the negative set assumed to be present under ordinary procedures. The effect was, however, significant at only about the 8% level. Despite this marginal level of significance, the results might be taken to suggest that knowledge of conditioning improves performance. There is evidence which indicates, however, that the observed effect may not be due to knowledge per se, but rather to the specific instructions and procedures used in this experiment. Spence and Taylor (**5**) report that knowledge obtained from course lectures given just prior to an experiment lead to an increase in the number of voluntary conditioners while no *S* run in the present experiment was so classified. The discarded *Ss*, who obtained their knowledge from a variety of sources, tended to perform like naive *Ss*, but the few cases involved make any conclusion tenuous. Therefore, on the basis of the present information, it seems possible that knowledge about conditioning obtained from different sources might affect conditioning differentially. Further research is needed for clarification. In any event, it would seem wise to insure homogeneity among *Ss* with respect to their information about conditioning by using knowledge-giving instructions. The present results also suggest that the precision of conditioning experiments may be increased by controlling for the sex of the *Ss*.

SUMMARY

Four groups of *Ss* differing with respect to sex and the type of instructions given were run 70 trials in an eyelid-conditioning experiment.

One group of each sex was instructed as to the nature of classical eyelid conditioning while the other group was not given this information. It was found that the females conditioned better than the males, and the knowledge-of-conditioning *Ss* conditioned better than the no-knowledge *Ss*. Both factors were, however, significant at only about the 8% level.

References

1. Grant, D. A. The influence of attitude on the conditioned eyelid response. *J. exp. Psychol.*, 1939, **25,** 333–346.
2. Hilgard, E. R. An algebraic analysis of conditioned discrimination in man. *Psychol. Rev.*, 1938, **45,** 472–496.
3. Hilgard, E. R., & Marquis, D. G. *Conditioning and learning.* New York: Appleton-Century, 1940.
4. Spence, K. W. *Behavior theory and conditioning.* New Haven: Yale Univer. Press, 1956.
5. Spence, K. W., & Taylor, J. Anxiety and strength of the UCS as determiners of the amount of eyelid conditioning. *J. exp Psychol.*, 1951, **42,** 183–188.

(Received July 15, 1957)

Journal of Experimental Psychology
Vol. 55, No. 6, 1958

HABIT REVERSAL AS A FUNCTION OF SCHEDULE OF REINFORCEMENT AND DRIVE STRENGTH[1]

HOWARD H. KENDLER AND ROY LACHMAN

New York University

It has been demonstrated that experimental extinction is retarded by intermittent reinforcements during acquisition (**7**) and high drive during extinction (**12**). If these findings reflect basic psychological principles that are applicable to experimental situations more complex than conditioning, then it would be expected that both the schedule of reinforcement and drive strength should influence a habit reversal in a simple discrimination situation. The reasoning behind this prediction is as follows: The *S*, in order to solve a discrimination problem, must acquire a stronger habit to approach the positive stimulus than to approach the negative stimulus. If, after solving the original discrimination, *S* is required to execute a reversal shift (i.e., learn to reverse his preference), then the approach response to the original positive stimulus will have to be weakened below the strength of the approach response to the original negative stimulus. Any factor that would retard this weakening process should also retard a habit reversal.

Experiments have been reported that bear upon our analysis. Wike (**16**) and Grosslight, Hall, and Scott (**4**) have demonstrated that in a nonspatial discrimination situation a habit reversal was executed more rapidly following the learning of the original habit under a continuous reinforcement schedule as contrasted with an intermittent reinforcement schedule. These results are, of course, consistent with the present analysis.

Of three recent studies (**2, 3, 11**) designed to investigate the influence of drive strength on habit reversal, none reported any retarding effect of high drive during reversal training. It need be mentioned, however, that in all three studies the habit reversal was executed fairly rapidly. It is possible that the retarding effect of high drive was hidden by this factor. The reason for believing this is that a habit reversal involves two major successive stages of learning. During the first stage of reversal learning the dominant habit is weakened. During the second stage the to-be-learned response, which previously was nonreinforced, is strengthened. Only during the first stage of the reversal learning would the high drive retard the reversal shift. Once the to-be-learned response (the originally negative habit) becomes dominant, then a high drive should facilitate a habit reversal (**14**).

It is necessary, therefore, when attempting to discover the influence of drive strength during habit reversal, to separate the successive interfering and facilitating effects of high drive during the aforementioned suc-

[1] This study was supported by the Office of Naval Research. It was part of a series of studies designed to investigate the retarding influence of habit persistence on problem solution. Reproduction in whole or in part of this article is permitted for any purpose of the United States Government. The authors wish to express their appreciation to M. R. D'Amato for his invaluable assistance in the design and construction of the apparatus.

cessive stages of reversal training.[2] This was done in the present study by deliberately designing an experimental problem that would produce slow reversal learning.

Method

Experimental design.—A 2 × 2 factorial design was adopted. During the original discrimination training half of the *Ss* (Group C) were trained under conditions of continuous reinforcement, while the remaining *Ss* (Group I) had an intermittent reinforcement schedule for their responses to the positive stimulus. During reversal learning half of the *Ss* in Groups I and C were motivated by a high drive (Groups IH and CH) while the remaining *Ss* in each group had a low drive (Groups CL and IL).

Apparatus.—The discrimination apparatus consisted of a Y maze possessing three identical arms. This arrangement permitted a continuous discrimination apparatus which dispensed with *E*'s handling of *S* between successive trials of the same day. The *S* was placed in one of the three arms which served, for that trial, as a starting compartment. One of the remaining two arms contained the positive stimulus, while the other one contained the negative stimulus. After *S* made his choice and completed the trial, the arm which he chose became the starting compartment for the next trial. The other two arms contained either the positive or negative stimulus. This procedure was continued until *S* completed his daily runs.

Each arm was spaced at 120° from each of the other two arms. The walls of the arms were constructed of .5-in. plywood painted gray. The floor, except for the rear 2 in. of each arm, consisted of unpainted .25-in. aluminum rods. The rear 2 in. consisted of a slidable wooden panel containing a circular depression (.25 in. deep and .5 in. in diameter) that was used as a food container. This wooden panel was removable.

The internal dimensions of each arm were as follows: 14 in. long, 4.5 in. wide, and 5 in. high. The entrance to each arm (3 × 3 in.) contained a door which was opened by lowering it to the level of the floor. When the door was raised, thus closing the entrance to the arm, a .25-in. space remained, permitting *S*'s tail to protrude. The top of each arm contained an observation screen 8 in. long by 2.5 in. wide, consisting of .5-in. hardware mesh. Each arm was illuminated by two 7-w. bulbs, one of which was located within the rear wall of the arm and was shielded by a 4.5 × 5-in. sheet of laminated fiberglass. The other bulb was located in the roof of the arm at the door and was shielded by a smaller sheet (2 × 4.5 in.) of the same material. Three conditions of illumination were possible: dark, light, and flickering light. When the illumination condition was *light*, the light intensity with the light meter facing into the arm was 8 to 10 ft.-candles; inside the arm at the door with the meter facing up the light intensity was 21 to 60 ft.-candles, while with the meter facing down the light intensity was .8 to 1.9 ft.-candles. The light intensity at the food cup was 50 to 80 ft.-candles when facing the rear illuminated wall. Facing down the light intensity was 5 to 16 ft.-candles. The light intensity of the *dark* arm was too low to be recorded with the meter used (Weston, Model 603).

Each arm functioned as follows: When the door was lowered, thus permitting entrance into the arm, an electrical timer was automatically started. When *S* penetrated the arm for a distance of at least 6 in., a low amperage contact relay circuit between *S*'s feet and the floor rods was completed, resulting in the clock being stopped.

Preliminary training.—The *Ss* were placed on a 24-hr. feeding schedule 7 days prior to the start of discrimination training. Purina laboratory chow pellets were available for 1.25 hr. daily. On Day 5 of preliminary training, a dozen .065-gm. reward pellets made from the Purina laboratory chow were fed to *Ss* prior to their regular daily feedings. On Days 6 and 7, immediately after their feeding period, each *S* was given a 15-min. exploratory session in groups of three or four. Each of the three conditions of illumination was presented in all three arms simultaneously for periods of 5 min. The food cup and rear section of the arms were covered with .065-gm. reward pellets. The *Ss*

[2] Eisman, Asimow, and Maltzman (**3**), who found that habit reversal was uninfluenced by drive strength, attributed their results to the successive interfering and facilitating effects of high drive during reversal learning. It is likely this same factor operated in Meyer's study (**11**). Buchwald and Yamaguchi (**2**), who found high drive facilitates reversal learning, used a correction procedure in a simple learning task (T maze) with a low criterion of learning (10 out of 12 correct responses). It is likely that as a result of their procedure the difference between the competing habits of *Ss* was very small at the termination of the original learning. If this were so then their results would *not* reflect that stage of reversal learning that primarily involves the weakening of the previously dominant habit.

consumed these pellets despite being under virtually zero drive.

Discrimination training.—The *S*s were randomly divided into two groups: Group I (intermittent reinforcements) and Group C (continuous reinforcements). Each *S* received 12 trials daily for a period of 13 days. The 12 daily trials consisted of four replicated blocks of three trials each. A block of three trials consisted of one free trial followed by two forced trials. The first forced trial involved forcing *S* to the stimulus opposite that chosen on the free trial. The second forced trial was always to the positive stimulus. This procedure resulted in all *S*s receiving daily eight trials to the positive stimulus and four trials to the negative stimulus. It should be noted in the present study, unlike previous studies (**2, 3, 12**) investigating the influence of levels of drive on reversal learning, that the number of responses to the original positive and negative stimuli were held constant for all *S*s. The forced trial was arranged by closing the door to one of the two arms. The *S*s in Group C were always rewarded when going into the arm containing the positive stimulus. The *S*s in Group I received a reward pellet on only some of their trials to the positive stimulus. On Day 1 *S*s in Group I received seven reinforced and one nonreinforced trial to the positive stimulus. On Day 2 the number of reinforced trials was reduced to six and on the following day to five. From Days 4 to 13, Group I *S*s received four reinforced and four nonreinforced trials on their eight trials to the positive stimulus. These nonreinforced trials to the positive stimulus were assigned in a random manner (regardless of whether they were free or forced trials) except for the following three restrictions: (*a*) The first daily response to the positive stimulus was always reinforced; (*b*) no more than three nonreinforced trials to the positive stimulus could occur consecutively, and (*c*) if a given free trial was scheduled for a nonreinforcement for a response to the positive stimulus and *S* chose the negative stimulus, then *S* was given a nonreinforcement on the subsequent forced trial. The food cup was present in the arm containing the positive stimulus on both reinforced and nonreinforced trials but it was always absent from the arm containing the negative stimulus.

During this initial discrimination training, the positive stimulus was an illuminated arm, the negative stimulus was a darkened arm. The location of the positive stimulus was randomly varied from the arms on *S*'s left or right on both free and forced trials. The intertrial interval on successive trials was approximately 30 sec. When *S* completed eating and/or the apparatus had been programmed for the following trial, the positive or negative stimulus of the arm containing *S* was replaced by a flickering light which lasted from 3 to 5 sec. and was followed by the darkening of the arm, the opening of its door, and the activation of a timing mechanism.

At the start of a day's training schedule, *S* was removed from the home cage in an adjoining room after being deprived of food for 22 hr. The *S* was then placed in a dark arm and soon thereafter the flickering light was presented.

On the first day of the 13-day discrimination series, *S* received three reward pellets for the first response to the positive arm. On the next response to the positive stimulus the number of food pellets was reduced to two. Thereafter the reinforcement was one pellet. Sixty seconds after the last daily trial (Trial 12), *S* was returned to its home cage and 15 to 20 min. later its Purina laboratory chow was placed in its cage for 1.25 hr. Water was always available in the home cage.

Reversal training.—The reversal series consisted of 12 daily free-choice trials for a period of 10 days. Choices to the dark arm (previous negative stimulus) were consistently reinforced, while choices to the light arm (previous positive stimulus) were consistently nonreinforced.

Prior to the start of each trial, the flickering light was flashed for a period of from 3 to 5 sec. Then the door was opened and the starting arm was darkened. If *S* retraced, the flickering light was reactivated to prevent *S* from confusing the starting box with the positive arm. Very little retracing, however, was noticed. On Day 1 of reversal training, three *S*s failed to discover the food pellets. The *E* assisted *S* by placing its head next to the food cup. The *S*s then consumed the food.

During the reversal training *S*s in Groups I and C were randomly subdivided into a high (H) and low (L) drive group. This resulted in four groups during reversal training: IH, IL, CH, and CL. All *S*s were fed for a 2-hr. period on alternate days. The *S*s in the H condition were run 45 hr. after the termination of their last feeding period, while *S*s in the L condition received their training trials 3 hr. after their feeding session. Because of this severe deprivation schedule the Purina laboratory chow of all *S*s was supplemented by an oil solution consisting of 80% peanut oil and 20% cod liver oil. One part oil (by volume) was mixed with 20 parts ground Purina laboratory chow. After this was done 12 parts of water were then added.

The intertrial interval was approximately 20 sec. The *S*s that made 12 correct choices for a given day were eliminated and were assigned a score of 12 correct trials for each of the days remaining in the reversal training series.

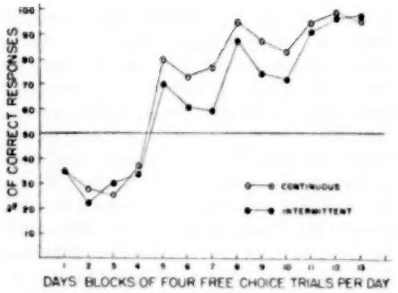

FIG. 1. The percentage of correct responses on free-choice trials for each of the 13 days during the initial discrimination learning.

RESULTS

Figure 1 shows the learning curves of Groups I and C during the initial discrimination learning. The curves begin at a below-chance level because of *Ss*' preference for the negative stimulus (dark). Both groups were responding at virtually a 100% level during the terminal portion of the training procedure. The performances of Groups C and I, however, were found to be different when compared by the Kolmogrov-Smirnov Test (16). A comparison of the distribution of each *S*'s total number of correct responses during the 52 free-choice trials revealed the two groups to be significantly different beyond the .01 level.

Figure 2 shows the reversal learning curves, while Table 1 reports the analysis of variance of correct responses following the reversal shift. The analysis produced three significant effects: schedule of reinforcement during the original discrimination, drive strength during reversal learning, and days during reversal discrimination learning. No significant interaction was obtained. Figure 2 graphically reveals the reason for the three significant effects. Figure 2 shows that the two intermittently reinforced groups (Groups IH and IL) had inferior performances during reversal learning when compared to their comparable motivated continuously reinforced groups (Groups CH and CL). Figure 2 also shows that the highly motivated groups (Groups CH and IH) performed at a lower level of proficiency during reversal learning than their comparable weakly motivated groups (Groups CL and IL). Thus these data show that intermittent rein-

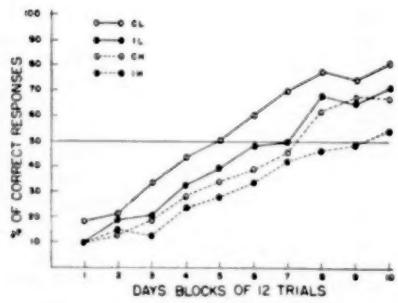

FIG. 2. The percentage of correct responses for each of the 10 days (12 trials per day) of reversal learning.

TABLE 1

ANALYSIS OF VARIANCE OF CORRECT RESPONSES FOLLOWING THE REVERSAL SHIFT

Source	*df*	*MS*	*F*
Between *Ss*	(47)		
Reinforcement schedule (R)	1	134.408	5.77*
Drive strength (DS)	1	285.208	12.23**
R × DS	1	6.075	
Error	44	23.246	
Within *Ss*	(432)		
Days (D)	9	309.008	130.383**
D × R	9	2.917	1.230
D × DS	9	4.013	1.693
D × DS × R	9	2.741	1.156
Error	396	2.370	
Total	479		

* $P = .05$.
** $P = .01$.

forcement during initial discrimination learning and high drive during reversal learning retard habit reversal. The third significant effects (Days) is due to the gradual increase in the number of correct choices for all four groups throughout the reversal learning period.

An analysis of the tendency to perseverate on the originally correct response following the reversal shift failed to yield a significant difference among the four groups. The mean number of choices made by each group to the formerly correct side (light) before making a run to the dark side was 14.8, 8.8, 8.8, and 5.6 for Groups IH, IL, CH, and CL, respectively. Although the pattern of results was similar to those reported in Fig. 2, a Krushal-Wallis one-way analysis of variance (**16**) by ranks failed to provide a significant difference.

Discussion

The main results obtained are consistent with our theoretical expectations. Both an intermittent reinforcement schedule during acquisition and high drive following the reversal shift retarded a habit reversal.

Although these results were clear-cut, certain explanatory problems remain unresolved. One such problem concerns the effect of high motivation during the second stage of the reversal shift. Once the correct postreversal shift response gains dominance, then a high drive should facilitate a habit reversal. According to such an analysis our statistical treatment should have revealed an interaction effect between Drive level and Days; i.e., after the performance curves of the high-drive group passed the 50% level they should have started to "catch up" with the low-drive groups.

The statistical analysis failed to reveal any interaction effects. This may be due in part to the fact that our experiment had to be terminated after 10 days of postreversal training. As a result, none of the four groups had all their *S*s reach a criterion of reversal learning. Although Group CH appeared (see Fig. 2) to be closing the gap between its performance curve and that of Group CL from Day 7 to Day 9, the trend failed to continue into Day 10. Group IH barely reached the 50% level by the end of the reversal-learning period and therefore had no opportunity to demonstrate any beneficial effect of a high drive.

A complicating feature of this problem is that *S*s do not shift directly from the originally correct discrimination response to the correct reversal response. Instead *S* usually passes through a stage in which he adopts a spatial response habit, i.e., *S* consistently chooses the right or left alley regardless of the cue.

Figure 3 reports the percentage of *S*s in each group responding to the same side for 10 or more of the 12 daily trials during reversal learning. Since all *S*s persist in their originally correct habit, the curves of each of the four groups start at 0%. The curve of Group IH, the group with the poorest performance during reversal learning, rises to the highest point in Fig. 3 and at the final day of the reversal-learning period has more than 50% of its *S*s responding in a position preference. Group CL, the group with the best performance during

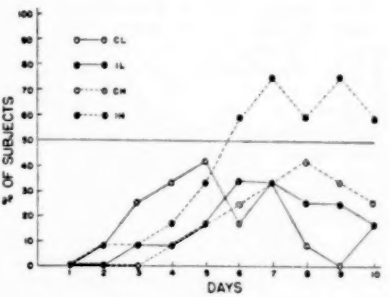

Fig. 3. Percentage of *S*s responding to the same side for 10 or more daily trials during reversal learning.

reversal learning, during the initial period of reversal learning has more *S*s responding in a position preference manner. However, this tendency begins to decrease after Day 5 and practically disappears by the end of the reversal-learning period. It would appear that reversal learning in the present experimental situation is initially blocked by the originally correct habit and then by position habits.

Consequently, our analysis of reversal learning into two successive stages involving the extinction of the initially learned response and then the strengthening of the to-be-learned response is an oversimplification. The transition from the originally correct habit to the correct reversal-learning response may pass through the successive extinctions of several other habits.[3]

Theoretically a high drive should impair performance as long as any incorrect habit is stronger than the correct habit. The high drive should retard extinction of all incorrect habits. Figure 3 suggests that Group IH takes longer to extinguish its position habits than Group IL, but there appears to be no difference in the performances of Groups CH and CL.

The second unresolved theoretical problem concerns the influence of drive stimulus generalization on our results. It is possible that the retarding influence of a high drive during reversal learning may be largely due, if not exclusively, to the fact that the high-drive group (45 hr.) suffered a smaller drive stimulus generalization decrement than the low-drive group (3 hr). That is, a drive stimulus based on 45 hr. of deprivation is more similar to a drive stimulus resulting from 22 hr. of deprivation than is a drive stimulus based on 3 hr. of deprivation. There are two reasons we believe this is unlikely. If a marked drive stimulus generalization effect operated, it should have been revealed on the first day of training. Groups IL and IH chose the originally positive stimulus an equal number of times. Although Group CL chose the originally positive stimulus a smaller percentage of times than did Group CH, the difference was not significant. The second reason is that available evidence suggests a very flat drive stimulus generalization gradient. Bloomberg and Webb (1) trained rats to make a right turning response in a T maze when motivated by 22 hr. of hunger and a left turning response when motivated by a 3-hr. hunger drive. In one experiment only one of 12 *S*s succeeded in learning the problem during a training period consisting of 130 reinforcements of each response. Bloomberg and Webb modified their experimental procedure in several ways in the hope of facilitating learning. In their second study 10 of 12 *S*s succeeded in reaching a criterion of learning only after extensive training (ranging from 115 to 185 reinforcements for *each* of the two training responses). Jenkins and Hanratty (6) were able to train five rats to make different spatial responses to cues resulting from 11½ hr. and 47½ hr. of food deprivation. Definite signs of learning did not occur until after 40 days of training (3 trials per day). After 50 days of training, at which time the experiment was terminated, *S*s still were unable to respond appropriately 100% of the time. If the drive stimulus generalization curves dropped sharply, learning would have occurred much more rapidly and effectively than it did in the two studies cited.

It might also be mentioned in passing that our state of knowledge concerning drive stimulus generalization is woefully inadequate. No parametric studies have been reported showing drive stimulus generalization curves. We are even ignorant about the identity of the receptor mechanisms underlying a drive stimulus. If differences in learning are to be attributed to drive stimulus generalization decrement effects, then there is an obvious need for basic

[3] Our analysis is oversimplified in one other respect. In the present experimental situation *S*s do not practice only one habit (e.g., position or brightness) at a time. A choice on any one trial will be influenced both by the position habit to one side as well as the approach habit to the brightness cue of the same side.

empirical evidence about the nature of the generalization curves.

An obvious question that might be raised is whether a more appropriate design for the problem we sought to investigate would not have involved a factorial design with two levels of drive strength during acquisition as well as during reversal learning. Such a design has been used frequently in studies investigating the role of drive on learning and performance (e.g., **8**). This experimental design was rejected in favor of the one used because it was desired that prior to the reversal shift all *S*s not only have the same amount of learning but also learn the same habits. Although different levels of drive during instrumental conditioning do not play a role in habit strength (**8, 15**), there is evidence that drive differences during the learning of a black-white discrimination does play a role in habit strength (**3**).

The relationship between drive and discrimination learning is an exceedingly complex one (**5, 14**). Different levels of drive, by producing different running speeds, would influence such a variable as delay of reinforcement, which in turn would affect the anticipatory goal response. Differences in the anticipatory goal responses could produce both drive and habit differences (**9, 10**). Consequently, if a factorial design were used it would be exceedingly difficult to unravel the separate influences of habit and drive differences resulting from different levels of drive during the original discrimination learning.

It would appear that no single experimental design is capable of yielding a definitive answer to the theoretical question posed. Both the factorial design and the one used in the present paper suffer from certain limitations. The reason for this state of affairs is that the sharp theoretical distinction between drive and habit is not mirrored in experimental manipulations. The variation of a single independent variable, such as deprivation period, can influence both drive and habit. As long as this state of affairs remains, several experiments, each with a different design, may be required to answer one theoretical question. Another alternative would be to isolate the empirical variables underlying the habit-drive distinction and thus make it possible to answer a single theoretical question with a single experiment.

Summary

Forty-eight rats learned a simple nonspatial discrimination (Phase 1). Half of the *S*s had intermittent reinforcements, while the other *S*s were continuously reinforced. During Phase 2 all *S*s were subjected to a reversal shift under a strong or weak drive. The results indicated that the habit reversal was retarded both by intermittent reinforcements during Phase 1 and the high drive during Phase 2.

References

1. Bloomberg, R., & Webb, W. B. Various degrees within a single drive as cues for spatial response learning in the white rat. *J. exp. Psychol.*, 1949, **39,** 628–637.
2. Buchwald, A. M., & Yamaguchi, H. G. The effect of change in drive level on habit reversal. *J. exp. Psychol.*, 1955, **50,** 265–269.
3. Eisman, E., Asimow, A., & Maltzman, I. Habit strength as a function of drive in a brightness discrimination problem. *J. exp. Psychol.*, 1956, **52,** 58–64.
4. Grosslight, J. H., Hall, J. F., & Scott, W. Reinforcement schedule in habit reversal —a confirmation. *J. exp. Psychol.*, 1954, **48,** 173–174.
5. Hull, C. L. *A behavior system.* New Haven: Yale Univer. Press, 1952.
6. Jenkins, J. J., & Hanratty, J. A. Drive intensity discrimination in the albino rat. *J. comp. physiol. Psychol.*, 1949, **42,** 228–232.
7. Jenkins, W. O., & Stanley, J. C. Partial reinforcement: A review and critique. *Psychol. Bull.*, 1950, **47,** 193–234.
8. Kendler, H. H. Drive interaction: II. Experimental analysis of the role of drive in learning theory. *J. exp. Psychol.*, 1945, **35,** 188–198.
9. Kendler, H. H. Some comments on O'Kelly and Heyer's interpretation of their experimental results. *J. comp. physiol. Psychol.*, 1949, **42,** 233–235.

10. Logan, F. A. A micromolar approach to behavior theory. *Psychol. Rev.*, 1956, **63,** 63–73.
11. Meyer, D. R. Food deprivation and discrimination reversal learning by monkeys. *J. exp. Psychol.*, 1951, **41,** 10–16.
12. Perin, C. T. Behavior potentiality as a joint function of the amount of training and the degree of hunger at the time of extinction. *J. exp. Psychol.*, 1942, **30,** 93–113.
13. Siegel, S. *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill, 1956.
14. Spence, K. W. *Behavior theory and conditioning.* New Haven: Yale Univer. Press, 1956.
15. Strassburger, R. C. Resistance to extinction of a conditioned operant as related to drive level of reinforcement. *J. exp. Psychol.*, 1950, **40,** 473–487.
16. Wike, E. L. Extinction of a partially and continuously reinforced response with and without a rewarded alternative. *J. exp. Psychol.*, 1953, **46,** 255–260.

(Received July 22, 1957)

Journal of Experimental Psychology
Vol. 55, No. 6, 1958

COMPARING REVERSAL AND NONREVERSAL SHIFTS IN CONCEPT FORMATION WITH PARTIAL REINFORCEMENT CONTROLLED [1]

MARTIN HARROW AND GILBERT B. FRIEDMAN [2]

Indiana University

Kendler and Vineberg (7) sought to determine the influence the learning of simple concepts had upon the learning of compound concepts, using a card-sorting task. These compound concepts involved combinations of simple concepts which had just been learned. They confirmed their expectation that the rate of learning the test concept would be directly related to the number of relevant simple concepts *S*s had previously learned. Further analysis suggested to them that the advantage derived from the acquisition of the appropriate simple concepts did not stem from the mere repetition of correct sorting responses. Rather, it came from the opportunity the initial training provided for the appropriate symbolic or verbal responses to become dominant and thereby facilitate the acquisition of the correct test responses. Thus, a mediational hypothesis which depended on implicit verbal responses was formulated. According to this formulation the card-sorting behavior on any one trial consisted of a sequence of two successive S-R associations. The stimulus component of the first association would represent the test cards while the response would refer to implicit verbal or symbolic responses made to them. The stimulus of the second association would represent the cue produced by the preceding implicit response, while the response would be the overt card-sorting behavior.

At about the same time, Buss (2) performed an experiment in concept formation comparing reversal shifts with nonreversal shifts. A reversal shift involves learning one discrimination, and then learning a second discrimination which is opposite in direction to the first. For example, in Discrimination 1, the color green is the positive stimulus and red is the negative stimulus, and in Discrimination 2, red is positive and green is negative. Nonreversal shifts involve a change in the dimension of the stimulus being discriminated. For example, in Discrimination 1 the dimension on which sorting is reinforced is color—green or red—and in Discrimination 2 the dimension on which discrimination is based is shape—a star or a cross. Buss' results showed that the reversal group learned quicker than did the nonreversal group. He accounted for these results in terms of partial reinforcement. That is, if the nonreversal *S*s continued to make the same responses during the learning of the second concept that they had previously been reinforced for during first concept learning, they would receive 50% reinforcement. The presence of this partial reinforcement could possibly slow up their shift and account for the superiority of the reversal group. Thus, either partial reinforcement of *S*'s responses, or a mediational hypothesis, could successfully account for the results of this experiment.

In order to resolve this issue, Kendler and D'Amato (6) carried out an experiment demonstrating the superiority of reversal over nonreversal shifts when

[1] The authors are greatly indebted to W. K. Estes and J. A. Dinsmoor for their generous advice and encouragement.

[2] Now at the University of California, Los Angeles.

partial reinforcement is eliminated. Their conclusion was that, since partial reinforcement had been eliminated, the results favored the mediational hypothesis. The specific applicability of this hypothesis was that the presence of the appropriate symbolic cues, even though they were connected to the opposite or wrong sorting response, would facilitate concept formation. However, their conclusions are based on the assumption that partial reinforcement was eliminated. An analysis of their design by Buss (**3**) led to the conclusion that partial reinforcement was not eliminated.

Buss (**3**) tried to eliminate partial reinforcement by modifying the Kendler and D'Amato experiment. However, the present writers believe that Buss' experiment also has several flaws. For one, his reversal and nonreversal groups (the comparison which forms the basis of the experiment) are neither randomized nor matched. The *S*s are broken up into these two groups according to their manner of responding to identical stimuli. Those *S*s who use the shape of the stimuli as the relevant cue to determine their sorting responses are put into one group, those who use color as the relevant cue for sorting are put into another group, and the two groups are then compared for speed of learning. Since the stimuli presented to all *S*s are the same, any difference in the type of response to these identical stimuli may be due to differences in *S*s. Under these conditions, with the groups neither randomized nor matched, and the question of differences between *S*s composing each group left unanswered, the results would seem somewhat ambiguous. Furthermore, Buss failed to control for any basic difference between the speed of learning shape and color concepts. His reversal group was composed of *S*s who used shape as the relevant cue to determine their sorting responses, while his nonreversals were *S*s who used color as the relevant cue for sorting. Buss justified this procedure by pointing out that a part of Kendler's experiment had failed to find any differences between the speed of learning reverse shape and color concepts. However, this point is at best inconclusive since the comparison that Buss speaks of (**6**, p. 168, Table 3) is actually one between two groups which are labeled differently (shape, and color) but have identical sorting tasks. They are given different labels because later on in the experiment, *after* the crucial comparison of interest has already been made, these two groups then have different tasks assigned to them. Thus Buss' comparison of the medians of two groups having identical tasks is somewhat inappropriate. Furthermore, Kendler and D'Amato also reported a significant difference between a shape group and a color group, in favor of the former (**6**, p. 168, Table 2). Hence, the possibility exists that on the second task Buss' reversal group, which learned a shape concept, learned quicker than his nonreversal group, which learned a color concept, simply because a shape concept is often learned quicker. Thus, there is at least some evidence that justifies the opinion that this type of control is one which warrants further caution before it can be eliminated. It was the purpose of the present experiment to compare reversal shifts with nonreversal shifts in concept formation, while using a more adequate and direct way of eliminating partial reinforcement of *S*s' previous sorting responses.

Method

Subjects.—The *S*s were 69 undergraduate students from introductory psychology courses at Indiana University. Two were eliminated for failure to learn the first concept, and the data from one other *S*, randomly selected, was discarded in order to equalize the *N*'s in groups.

Design.—The main purpose of the experiment was to compare the relative effectiveness of a reversal shift as opposed to a nonreversal shift, with partial reinforcement eliminated. This was accomplished in a conventional card-sorting task by a factorial design in which half the *S*s were in a reverse number group and the other half were in a reverse color group. In the reverse number group one third of the *S*s, the reversal *S*s, first learned to discriminate on the basis of number and then learned a reverse number concept. Another third, the nonreversal *S*s, first learned to

discriminate on the basis of color and then shifted to a reverse number concept. The last third, which were the control *Ss*, only learned the reverse number concept. Likewise with the reverse color group, one third, the reversals, first learned to discriminate on the basis of color and then learned a reverse color series. Another third, the nonreversals, first learned to discriminate on the basis of number and then shifted to a reverse color series. The remaining third, who were controls, learned only the reverse color series. Thus, six groups were formed: (1a) number-reverse number; (1b) color-reverse number; (1c) nothing-reverse number; (2a) color-reverse color; (2b) number-reverse color; and (2c) nothing-reverse color.

Partial reinforcement of the first concept responses, during *Ss*' learning of the second concept, was eliminated by using a different set of cards during this second stage of the experiment. Hence, all experimental *Ss* were presented with different sets of cards for the first and second concepts they learned. This permitted a situation in which *Ss* were never reinforced during the second stage of the experiment for a response they had previously been reinforced for during the learning of the first task. Thus the experimental *Ss* in the reverse-number group learned the first concept with a set of cards in which each card had one of four shapes represented on it—stars, circles, squares, or triangles—each card had either one or two identical figures on it, and the figures on each card were either red or green in color. The set of cards used during the second series were like those of the first, except that all the figures were yellow or blue rather than red or green. Thus, the nonreversal subgroup, which is the crucial subgroup that has presented the major source of difficulty in the past in the way of partial reinforcement, was always reinforced on the basis of red positive and green negative during the learning of the first concept. During the second concept, this group was reinforced with two figures being positive and one negative, without having an opportunity to respond to red or green cards any more, as these cards were no longer presented to them.

The set of cards presented to the reverse-color groups during the learning of the first concept was the same as those presented to the reverse-number groups. That is, each card had one of the four shapes represented on it, all were either red or green and all had one or two identical figures on them. During their learning of the second series, the cards presented to them were the same as those in the first series, except that they all had three or four identical figures on them instead of one or two. Thus, in a manner similar to that for the reverse-number groups, partial reinforcement was also eliminated for the reverse-color groups. A summary of the experimental design is presented in Table 1.

Apparatus and materials.—Cards from the Wisconsin Card Sorting Test (1) were used. The regular deck of 64 cards was divided into three smaller decks of 16 cards each, with the remaining 16 cards being discarded. Each card had figures on it which varied according to the form, number, and color of the figures. A fourth way that they varied was in the asymmetrical placement of the figures on the cards. Each set of 16 cards used four different forms—two different colors and two different numbers of figures on the cards.

A large screen-type apparatus was used which shielded *E* from *S*'s view. There were three slots in this apparatus through which the cards could be passed: on the right, in the middle, and on the left.

Procedure.—The response cards were passed to *Ss* through the center slot. The *Ss* were instructed to pass them back through the slot on either their right or left side. The slot on

TABLE 1

EXPERIMENTAL DESIGN

Groups	Series 1		Series 2	
Reverse Number	+	−	+	−
1a. reverse (N-RN)	R*1*, G*1*	R*2*, G*2*	Y*2*, B*2*	Y*1*, B*1*
1b. nonrev. (C-RN)	*R1*, *R2*	*G1*, *G2*	Y*2*, B*2*	Y*1*, B*1*
1c. control (O-RN)			Y*2*, B*2*	Y*1*, B*1*
Reverse Color	+	−	+	−
2a. reverse (C-RC)	*R1*, *R2*	*G1*, *G2*	*G3*, *G4*	*R3*, *R4*
2b. nonrev. (N-RC)	R*1*, G*1*	R*2*, G*2*	*G3*, *G4*	*R3*, *R4*
2c. control (O-RC)			*G3*, *G4*	*R3*, *R4*

Note.—Y1 = cards with one yellow figure; B2 = cards with two blue figures; R3 = cards with three red figures; and G4 = cards with four green figures. Relevant concepts are in italics.

S's left was arbitrarily designated by *E* as the side for positive stimuli and the slot on *S*'s right side as the one for negative stimuli.

The instructions read to *S* at the beginning of the experiment were: "You will be shown a series of cards, one at a time. Some of these cards belong in this slot [pointing] and some belong in this slot [pointing]. The idea is for you to find out which one they belong in. You must put each one in one slot or the other. Decide as best you can. If you are correct, I will say 'correct,' if not, I'll say 'incorrect.' Are there any questions?"

The change in the pattern of reinforcement and in the set of cards used for the second concept, was done without informing *S*s. They were allowed a limit of 96 trials, or six decks of 16 cards, in order to learn the first concept. Once the first concept had been learned, they received as many trials as was necessary for them to learn the second one. The criterion of learning for both concepts was eight successive correct responses.

Results

The results on the learning of the first concept are presented in Table 2. These scores represent the number of trials to learn the task, excluding the criterion trials. An analysis of variance was carried out on these data in order to determine whether the four groups were equated on their initial tasks. Due to the skewness of the data and the small value of some of the means of the groups, a modification of the square-root transformation ($\sqrt{X + .5}$) was used in place of the raw scores to obtain homogeneity of variance (**4**). The analysis showed no significant differences among the four experimental groups who learned the first tasks ($P < .05$). The over-all F for the four groups was 1.28, with 3 and 40 *df*.

TABLE 2

Number of Trials to Learn First Concept

Group	*N*	Mean	*SD*	Mdn.	Range
N-RN	11	12.3	14.5	5	0–38
C-RN	11	13.2	17.6	4	0–55
C-RC	11	12.5	12.6	12	1–39
N-RC	11	16.5	20.6	16	0–70

TABLE 3

Number of Trials to Learn Second Concept

Group	*N*	Mean	*SD*	Mdn.	Range
N-RN	11	3.9	5.7	2	1–20
C-RN	11	14.4	9.4	11	0–29
O-RN	11	12.5	18.5	5	0–62
C-RC	11	8.8	14.5	3	1–44
N-RC	11	24.7	22.4	21	3–70
O-RC	11	4.4	5.7	4	0–20

The data concerned with the learning of the second concept are presented in Table 3. In testing for differences among the speed of learning these concepts, a 2×3 analysis of variance was performed, again using the same transformation to obtain homogeneity of variance. The results as shown in Table 4, of learning a number concept as opposed to learning a color concept, were nonsignificant with an F of .90. The F obtained for the over-all differences between the major experimental treatments, that between the reversal, nonreversal, and control groups, was 8.73 ($P < .001$). A further breakdown of this comparison was made by individual t tests, using the transformed scores and the mean

TABLE 4

Analysis of Variance of Trials to Learn Second Concept*

Source	*df*	*MS*	*F*
Treatments (T)	2	23.84	8.73**
Concepts (C)	1	2.63	0.96
T × C	2	5.63	2.06
Within groups	60	2.73	
Total	65		

* Transformed scores ($\sqrt{X + .5}$) used.
** $P < .001$.

TABLE 5

INDIVIDUAL *t* TESTS OF TRIALS TO LEARN SECOND CONCEPT

Comparisons	*t*	*P*
Reversals vs. Nonreversals	3.88	<.001
Reversals vs. Controls	.60	
Nonreversals vs. Controls	3.28	<.01
Color-Rev. No. vs. No.-Rev. Color	1.34	
No.-Rev. No. vs. Color-Rev. No.	2.46	<.02
Color-Rev. Color vs. No.-Rev. Color	2.98	<.01

Note.—Transformed scores ($\sqrt{X} + .5$) were used, with 60 *df* used for each *t* test.

square within groups to obtain the error term. These can be seen in Table 5. The test between the means of the reversal and nonreversal groups gave a significant *t* of 3.88 with 60 *df* ($P < .001$). Thus, with partial reinforcement eliminated from the experimental design, the reversal groups learned significantly faster than the nonreversal groups. The test for negative transfer between the nonreversal and control groups was also significant ($t = 3.28$, 60 *df*, $P < .01$). That between the reversal and control group, for positive transfer, was not significant ($t = .60$, 60 *df*).

The nonsignificant *F* of 2.06 for the interaction term in Table 4 indicates that the two concepts used in the second learning series (color and number) did not have a differential effect on the two nonreversal groups. Further evidence that there was no differential effect from the two different concepts used, was obtained when the reversal and nonreversal groups in Series 2 were broken down into their component groups and individual *t* tests were performed (Table 5). The test between the two nonreversal groups showed no significant differences ($t = 1.34$). A comparison between the individual reversal and nonreversal groups showed significant differences in both cases.

DISCUSSION

In the present experiment, *S*s were never reinforced during the reversal or nonreversal series for a response they had previously learned during the first series. Under these conditions, the superiority of the reversal over the nonreversal groups was significant at the .001 level.

It can be pointed out, in evaluating the two hypotheses so far presented, that one could not account for all the results by partial reinforcement. On the other hand, a mediational hypothesis based on implicit verbal or symbolic cues which Kendler has put into S-R language in the form of a double S-R sequence, is sufficient. However, it should be noted that thus far this hypothesis has been used explicitly (*a*) in accounting for the superiority of reversal shifts over nonreversal shifts when the initial concepts have already been acquired (**3, 6**), and (*b*) to account for the superior speed of learning complex concepts when the simple ones that compose it have already been acquired (**7**). In both types of situations the hypothesis is used explicitly to account for results which occur *after* one or a group of sorting responses have been acquired, and thus after the appropriate implicit verbal or symbolic responses connected with them have already become dominant. In both types of situations its use in explaining the acquisition of the first task before the reversal, or more complex, tasks are given is more or less implicit. If one were to be more explicit concerning its role in the acquisition of the first concept, one might say that *E*'s reinforcement of one of several sorting responses (which in themselves are usually based on cues from implicit verbal or symbolic responses that *S* has usually carried over from previous situations into the present one) leads to a new series of

implicit symbolic responses. The cues from these are used as the basis of sorting, and if incorrect, are progressively abandoned due to the resulting nonreinforcement of *S*'s sorting responses by *E*. Eventually the "correct" implicit response occurs. When overt responses to the cues produced by this one are always reinforced, they persist and are responsible for the correct placement of the remainder of the cards in the deck, even though *S* may not have seen these remaining cards previously. When this occurs, the implicit symbolic responses and the resulting cues from them would then persist and be responsible in many cases for the quicker learning of reversal concepts and slower learning of nonreversal concepts when the tasks are changed. The mediational process which forms the basis of this somewhat expanded S-R sequence was suggested by verbalizations of our *S*s after they had learned the tasks given to them.[3]

One might predict from the above type of mediational approach that, not only would a reversal shift be superior to a nonreversal shift to a different dimension but also, due to the resulting implicit verbal or symbolic cues, a nonreversal shift within the same dimension would be superior to a nonreversal shift to another dimension. This could not be tested in a conventional two-category situation, since in such a situation a change within the same dimension would necessitate the exact opposite sorting responses, a reversal-sorting task. However, it could be tested in a three- or multicategory situation, as here a change within the same dimension would not necessitate an exact opposite sorting response. A further prediction which follows from the above approach is that if *S*s' initial sorting responses are based on the cues from implicit responses carried into the situation from previous experience, then these sorting responses would not be random. Thus, in a situation where response cards must be sorted with one of several stimulus cards, *S*s would be expected to follow a definite pattern in their first few attempts at sorting.

[3] Further evidence in support of a mediational hypothesis is supplied in a recent experiment by Gormezano and Grant (**5**). In their experiment, using card-sorting tasks, partial reinforcement was given to an incorrect or irrelevant dimension (number) during *S*'s learning of a color concept. Later, during a second learning stage when the formerly irrelevant dimension (number) became relevant, there was no positive transfer as a result of the previous partial reinforcement. The authors conclude that these results are best accounted for by a mediational position.

Summary

Two previous experiments in concept formation have attempted to test the superiority of a reversal shift over a nonreversal shift in the sorting of cards and blocks, when partial reinforcement of the responses learned during solution of a first concept was not present during reinforcement of the second concept. Neither of these was completely successful. The present experiment eliminated partial reinforcement by using a different set of sorting cards for each of the two stages of the experiment. The *S*s were never reinforced during the learning of the second concept for a response they had previously learned in the first series. Two types of concepts were used: a number and a color concept. One set of reversal, nonreversal, and control groups learned a reverse number concept and another set of reversals, nonreversals, and controls learned a reverse color concept.

With partial reinforcement eliminated, the data from 66 *S*s indicated the superiority of the reversal groups over the nonreversal groups ($P = .001$). Evidence was also found of negative transfer for the nonreversal groups, when compared to control groups. The results were interpreted as indicating that partial reinforcement is not a necessary condition for explaining the superiority of reversal shifts over nonreversal shifts in concept formation. The evidence in support of a mediational formulation was discussed.

References

1. Berg, E. A. A simple objective technique for measuring flexibility in thinking. *J. gen. Psychol.*, 1948, **39**, 15–22.
2. Buss, A. H. Rigidity as a function of reversal and nonreversal shifts in the learning of successive discriminations. *J. exp. Psychol.*, 1953, **45**, 75–81.

3. Buss, A. H. Reversal and nonreversal shifts in concept formation with partial reinforcement eliminated. *J. exp. Psychol.*, 1956, **52**, 162–166.
4. Edwards, A. L. *Experimental design in psychological research.* New York: Rinehart, 1950.
5. Gormezano, I., & Grant, D. A. Progressive ambiguity in the attainment of concepts on the Wisconsin Card Sorting Test. *J. exp. Psychol.*, 1958, **55**, 621–627.
6. Kendler, H. H., & D'Amato, M. J. A comparison of reversal shifts and nonreversal shifts in human concept formation behavior. *J. exp. Psychol.*, 1955, **49**, 165–174.
7. Kendler, H. H. & Vineberg, R. The acquisition of compound concepts as a function of previous training. *J. exp. Psychol.*, 1954, **48**, 252–258.

(Received July 22, 1957)

Journal of Experimental Psychology
Vol. 55, No. 6, 1958

SUMMATION OF MANIFEST ANXIETY AND MUSCULAR TENSION [1]

DONALD R. MEYER AND MERRILL E. NOBLE

The Ohio State University *Kansas State College*

The study to be reported here is based upon the notion that the nonassociative properties of drives (**4, 8**) are due to interactions within efferent mechanisms. In this view, a drive is a response, and its capacity to alter the properties of other simultaneous responses stems from their mutual convergence upon the final common path. In essence, the interaction theory (**6**) is an extension of the principles which Sherrington deduced from his studies of reflex interaction and irradiation. Here the laws, instead of being limited in application to relatively localized and unlearned actions, are considered relevant to learned behaviors and to organismic patterns.

To be a generally effective drive, a response must be of such a form that its efferent excitation pattern is capable of changing the efferent networks for a large variety of actions. Only a strongly irradiative pattern has this property, which serves as the presumptive basis of the Hullian concept of generalized drive (**5**). Nonetheless, the theory holds that any response is capable of serving as a drive for at least some other responses, and that studies of performance as a function of drive are variants of dynamogenetic analyses.

In a recent study, Bourne (**2**) has shown that gripping a hand dynamometer has an effect upon recall in verbal learning which, in contrast with the earlier conception of Bills (**1**), can be purely nonassociative. Hence, the action of the hand, which is clearly a response, has a property of drive. The present experiment has been designed to see if such a drive will summate with conventional drives and produce an alteration in performance which would be expected from an increment in either. This is done by having groups of Ss who score low or high on the Taylor Manifest Anxiety Scale (**7**) learn a verbal maze under varying conditions of muscular induction

Both the pattern of "manifest anxiety" and gripping a dynamometer should alter verbal maze performance, and this should further vary as a function of the stage of practice. First, the theory holds that no response can be facilitated unless it has a certain habit strength and unless this habit strength is less than maximal. Within these limits, associative stimuli initiate a chain of events which lead to the capture of a portion of the efferent network for the pattern of behavior which has been selected for measurement. Input terminals for other portions of the network, which contemporary data indicate lie adjacent to ones already brought to threshold, will be biased catelectrotonically. Thus a simultaneous compatible response, if its

[1] This research was supported in part by the United States Air Force under Contract No. AF 33(038)-10528 with the Ohio State University Research Foundation, monitored by the Air Force Personnel and Training Research Center, and was conducted in the Laboratory of Aviation Psychology. Permission is granted for reproduction, publication, use, and disposal in whole or in part by or for the United States Government.

efferent excitation pattern is at least partially convergent with the excitation pattern induced by presentation of the stimulus, will selectively recruit those neurons whose operation will contribute to elicitation of the criterion performance. This is the presumptive basis of the dynamogenetic effect or, in other language, the energizing property of drive.

When, through learning, the stimulus-induced excitation comes to capture all of the neurons of an optimal network for a given response, further excitation derived from simultaneous compatible responses cannot yield facilitation; but, if such responses are equally compatible with actions antagonistic to responses which are chosen as correct, these antagonistic responses are facilitated. Such facilitation leads to a decrement in the criterion performance, a decrement that can be enhanced if extra-associative excitation is delivered to the networks at a level which transcends the point at which the optimal networks for criterion responses are fully recruited.

In the present situation, muscular induction and "manifest anxiety" are both assumed to be compatible with "right" and "left" as verbal responses. Further, the two driving responses are looked upon as mutually compatible. All three patterns, driving and criterion, are held to be mutually convergent. From these considerations, the prediction is that highly anxious persons are likely to do better than their nonanxious counterparts early in learning, if at all; if practice is continued to a high criterion, the latter subjects should excel. Added tension should be detrimental for the highly anxious *Ss*, either by way of eliminating their early superiority or by shortening the practice interval in which their superiority is manifest. To induce tension in *Ss* who score low in "manifest anxiety" should facilitate performance, but only if learning does not proceed to a high criterion. These effects, upon analysis, are expected to appear as a first-order interaction of tension and anxiety, and a second-order interaction of tension, anxiety, and practice.

Method

Subjects.—The *Ss* were 80 men and women from an introductory course in psychology. They were divided into groups of 40 on the basis of their scores on the modified Taylor Anxiety Scale (**7**). The raw scores for *Ss* in the high and low groups were 21–39 and 0–4, respectively; these are somewhat different ranges from the ones obtained by other workers, but together represent the deviant 30% of the sampled group. The high and low groups selected in this manner were subsequently subdivided, and 20 *Ss* in each learned with or without added tension.

Apparatus.—The task situation was arranged in a booth designed to standardize the visual situation. Within this booth, below a false bottom, there was placed a dynamometer with a centrally located vertical handgrip. Adjustments of the pressure required to close the grip could be made through manipulations of the mechanical advantage of a lever system linking the grip to a standard weight. The grip pressure was standardized at 10 lb. for all *Ss* working under tension. Limit switches on the grip controlled the lighting of the booth and thus provided a cue to *S* and *E* for proper maintenance of gripping.

The learning apparatus was placed within the booth atop the false bottom and consisted of two standard memory drums. One of these was used to present a verbal maze, the other to provide stimuli for color naming.

Procedure.—The procedure was largely a duplication of conditions which prevailed in an experiment of Taylor and Spence (**9, 10**). The principal change was interpolation of a 20-sec. interval between list presentations; this was done to minimize cumulative work effects in groups who learned under grip tension. The verbal maze was a 20-unit series of the words "left" and "right," these appeared sequentially in the windows of the first memory drum. Each word was exposed for 2-sec., and the task of *S* was to anticipate the next word in the series. After presentation of the list was complete, *Ss* in the grip-tension groups relaxed the handle. All *Ss* engaged in color naming between trials.

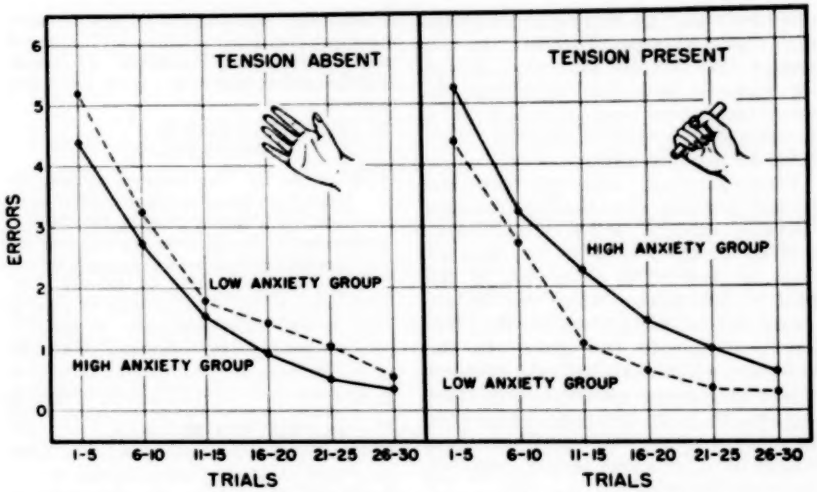

FIG. 1. The mean learning curves for groups which practiced with tension absent and tension present.

Thus a total cycle—learning and rest—was completed in 1 min. All *Ss* were given a total of 30 trials, regardless of the level of performance achieved. No score was taken on the color-naming task since it was used merely to keep *S* occupied between trials.

RESULTS AND DISCUSSION

The learning curves for all groups of *Ss* are shown in Fig. 1. Analysis of variance, a summary of which is given in Table 1, was used to test the differences. When tension is absent, the high group excels, although not significantly so; this relation is reversed by the induction of tension, although once again the difference within conditions is not significant. However, the interaction of tension and anxiety—which represents the change in the relative positions of the groups—is significant beyond the .05 level. This confirms our principal prediction.

The failure to show a substantial interaction of tension, anxiety, and practice presents a dilemma for the theory for it was also predicted that both inductions should eventually become a hindrance to maze performance. Why, then, does not the low-anxiety group eventually get better than the high under tension-absent conditions? One possibility is length of the practice session: Taylor and Spence (**9**), who utilized learning to criterion as a measure, found that their highly anxious *Ss* were significantly slower in the learning of

TABLE 1

ANALYSIS OF VARIANCE OF ERROR SCORES

Source	*df*	*MS*	*F*
Independent Obs.			
(T)	1	.96	—
(A)	1	2.33	—
TA	1	41.88	6.12*
Net between *Ss*	76	6.84	
Total between *Ss*	79	—	
Correlated Obs.			
Tr	5	216.86	309.80**
Tr × T	5	.22	—
Tr × A	5	.29	—
Tr × T × A	5	.58	—
Net within *Ss*	380	.70	
Total within *Ss*	400	—	

* $P < .05$.
** $P < .001$.

this very task. On this point, we prefer their results to ours. Together, the two sets of data indicate that what is true for muscular induction—that one obtains facilitation *and* inhibition as a function of the stage of practice (**3**)—holds for manifest anxiety as well.

Such parallels, together with the present demonstration that one can sum "manifest anxiety" and muscular induction to obtain results to be expected from an increase in either, are of the kind of information which will be required for an integration of the data on dynamogenesis and drive. In the literature which deals with the problem of how one response affects another, there is a wealth of information of relevance to modern drive theory. That these studies have been very largely forgotten is a minor tragedy.

Summary

This experiment was designed to ascertain whether two dissimilar responses—manifest anxiety and tension in the hand—can be summed to yield the effect to be expected from an increase in the magnitude of either. Forty volunteers from elementary courses in psychology, 20 rated high and 20 rated low in manifest anxiety, served as *Ss*. Half of the members of each of these groups learned a verbal maze while gripping a dynamometer; the remaining *Ss* served as a control.

The principal result was an interaction of tension and anxiety. The *Ss* in the low-tension or low-anxiety group tended to improve, while those with high scores deteriorated; the over-all effect was statistically significant. The results were discussed in relation to a motor interaction theory of drive.

References

1. Bills, A. G. The influence of muscular tension on the efficiency of mental work. *Amer. J. Psychol.*, 1927, **38,** 227–251.
2. Bourne, L. E., Jr. An evaluation of the effect of induced tension on performance. *J. exp. Psychol.*, 1955, **49,** 418–422.
3. Courts, F. A. The influence of practice on the dynamogenic effect of muscular tension. *J. exp. Psychol.*, 1942, **30,** 504–511.
4. Farber, I. E. The role of motivation in verbal learning and performance. *Psychol. Bull.*, 1955, **52,** 311–327.
5. Hull, C. L. *Principles of behavior.* New York: Appleton-Century, 1943.
6. Meyer, D. R. On the interaction of simultaneous responses. *Psychol. Bull.*, 1953, **50,** 204–220.
7. Taylor, J. A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48,** 285–290.
8. Taylor, J. A. Drive theory and manifest anxiety. *Psychol. Bull.*, 1956, **53,** 303–320.
9. Taylor, J. A., & Spence, K. W. The relationship of anxiety level to performance in serial learning. *J. exp. Psychol.*, 1952, **44,** 61–64.
10. Taylor, J. A., & Spence, K. W. Erratum. *J. exp. Psychol.*, 1953, **46,** 462.

(Received July 26, 1957)

Journal of Experimental Psychology
Vol. 55, No. 6, 1958

VARIATION OF TEMPORAL INTERVALS AMONG CRITICAL EVENTS IN FIVE STUDIES OF KNOWLEDGE OF RESULTS[1]

EDWARD A. BILODEAU AND INA McD. BILODEAU

Tulane University

The writers, having taken part in several unpublished experiments which have failed to show that delay of knowledge of results (KR) is detrimental to the learning of simple motor skill, and knowing of unsuccessful attempts by others, have been skeptical of the unqualified statement that delay of KR retards motor learning. Positive results to date are lacking when the delay interval is relatively free of interpolated activity. Positive results, however, have been obtained when the interval between responding and KR was filled by some similar activity. The studies reported here are of the former variety.

The three major temporal intervals studied, the ones which are common to all studies of KR delay, are identified in Fig. 1. Delay of KR represents the time between response completion and KR, the post-KR period represents the time between KR and the next response, and the intertrial interval represents the sum of the KR and post-KR delay periods. Figure 1 shows that an experimental design which varies delay of KR must either vary the post-KR period in the opposite direction in order to hold constant the value of intertrial interval, or hold the post-KR period constant and thus vary the intertrial interval. A question which arises is whether we are dealing with two or three variables. Experimentally, there are three manipulable variables, but knowledge of any two would suffice for purposes of psychological theorizing. Almost universal is the view that learning varies inversely with the duration of KR delay; an uncommon view is not that learning varies directly with delay of KR, but that learning varies inversely with the value of post-KR delay. Whichever experimental delay period is favored for purposes of theorizing, a position on the effects of intertrial interval should also be taken.

Reviews of the literature by Wolfle (**17**) and Ammons (**2**) would suggest that learning varies inversely with the duration of KR delay, and that no qualifications or special cases need to be made. Lorge and Thorndike (**8**), however, are skeptical about a retardation of learning as a function of increasing delay of KR, particularly where the delay period is unfilled and where it is possible to reinstate some equivalent of the response aftereffect. Brown (**5**) has observed that learning might in some cases vary inversely with the duration of the

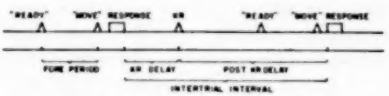

FIG. 1. A schematic representation of the sequence of events and the nomenclature of a typical study of knowledge of results.

[1] The raw data for the first four experiments were collected at the Air Force Personnel and Training Research Center. The assistance of Daniel Hitt in collecting the data of Exp. 5 is gratefully acknowledged. The preparation of this report was supported by funds of the University Council on Research of Tulane University.

post-KR period provided the response to be learned was relatively simple and discrete.

Uncritical generalizations of distribution of practice phenomena would suggest that performance improves with increasing values of the intertrial interval. If the response, however, is simple, not effortful, of brief duration, and only relatively massed, decremental states might not be important factors. Assuming that inhibitory states are negligible, forgetting interpretations predict that performance will vary inversely with the value of the intertrial interval. Indeed, the massed-practice data of Macpherson, Dees, and Grindley (**9**) are superior to their spaced-practice data.

Five related studies were designed to explore systematically the effects of the three major intervals. The responses taught were limited to relatively simple acts, and were required at discrete intervals. No similar activity was interpolated between the response and KR. The type of task used was similar to line drawing, and is to be contrasted with complex tasks such as tracking, maze, and questionnaire. Task simplification was thought a desirable criterion, for it was recalled that in a related area—delay of reward with animals—progress was related to the application of this criterion.

The five experiments varied each of the values of the three experimental variables through unusually wide ranges, employed uncommonly large numbers of *Ss*, and introduced a reminder technique as a control over potential effects of the post-KR period. The authors, having sided with Lorge and Thorndike, did not expect to find the acquisition of a relatively simple, discrete response retarded by delay of KR.

General Method

The *Ss* were 793 naive males, 180, 175, 300, 70, and 68 being assigned to Exp. 1 through 5, respectively. Except where noted the *Ss* were unsystematically assigned to their respective groups, and the treatments were administered in counterbalanced order.

Three highly similar tasks were used: knob turning, lever pulling, and stick displacing. All tasks were employed in the same way. The *S* was told that the assignment was to learn to make the optimal displacement of the control, for example: "Your task is to find out how far you have to pull this lever." After *S* displaced the control, he was given KR. This was to be used as a stimulus to improve the subsequent response. The first trial (Response 1) was made in the complete absence of information regarding the optimal extent of control displacement. The *S* responded while looking away from the control or while blindfolded. In all cases, *S* completed the response and returned his hand to a neutral or resting position before getting KR.

Experiment 1

The *S*'s task was to learn to turn a knob eight 360° turns. The experimental variable was the duration of KR delay measured in seconds.

Method

Apparatus.—The apparatus was a micrometer mounted so that the knob projected through a small hole in the center of a screen. The *S* was seated on the knob side of the screen; *E* was seated on the side of the micrometer scale. The extent of knob turning was measured by the micrometer, and read and recorded by *E*. After appropriate rounding and delay, *S* was given his score (verbally) by *E*. Each full turn of the knob accumulated 25 KR units. Thus, the goal was KR of 200, and eight 25-unit turns constituted the perfect response.

Procedure.—All *Ss* were told that the knob had to be turned to the point where a score of 200 would be reported by *E*, and that the idea of the test was to find out by means of the score reported after each trial what amount of turning was associated with a goal of 200. As a general aid to discovering the turning-score relationship, all *Ss* were required to make one full turn for Trial 1. After Trial 1, any number of turns was permitted.

A summary of the sequence and timing of within-trial events is given in Table 1. The major difference between the three groups was

TABLE 1

TIMING IN SECONDS OF WITHIN-TRIAL EVENTS FOR EXP. 1

Group	*N*	Fore-period	KR Delay	Re-minder*
5 sec. (R)	60	3	5	20
15 sec. (R)	60	3	15	20
20 sec.	60	3	20	—

* The reminder follows 20 sec. after the response. The ready signal follows 5 sec. after the reminder.

KR delayed by 5, 15, or 20 sec. The two shorter delay groups, however, had in addition a repetition or reminder (R) of KR 5 sec. before the ready signal. This procedure was introduced as a control over the effects of forgetting the KR. The accuracy of Groups 5 and 15 might, then, be expected to exceed that of Group 20 by virtue of the shorter delay *and* the KR reminder. Compared with the usual or standard experimental design, the present one was deliberately biased in favor of the shorter delay groups.

Excepting the time required to execute the response, the intertrial interval required 28 sec. Thus, the design holds the intertrial interval constant across all groups by concomitant variation of KR delay and post-KR delay; as the former increased, the latter decreased by an equal amount. The reminder control of post-KR delay leaves KR delay the only interval variable over Groups 5, 15 (8 sec. post-reminder) and 20 (8 sec. post-KR).

In all, 15 trials were administered, the first 10 as described above, the remainder as described below. After Response 10, without notice, the goal response was changed from 8 to 4 turns. This was done by adding 100 units to *S*'s score. For example, for 4 turns, KR becomes $4 \times 25 + 100$, or 200. Obviously, the last 5 trials bear a resemblance to extinction through response competition, and provide an additional test of the effect of delay of KR.

Results[2]

The mean absolute response errors for each of the 15 trials were computed, searched for significant differences between group means, and none was found.

[2] The data summarized in this report have been deposited with the American Documentation Institute. Order Document No. 5540, remitting in advance $1.25 for microfilm or $1.25 for photocopies.

Initially, the mean errors are large; by Trial 10 the errors are relatively small. On Trial 11, with a new response required, the mean error is again large, but by Trial 15 the error is much smaller. A between-group comparison of means shows some reversal of ranks from trial to trial. The longest delay group shows just a tendency to have the smallest error. Because the effect of differential delay might be expected to occur early in practice, the Trial 2 data were used in an analysis of variance: The *F* for between-groups was less than unity. Comparable analyses on the data of Trials 10 and 11 also yielded nonsignificant values of *F*.

Thus, despite shorter delays of KR *and* reminders of KR, two groups failed to differ significantly from a group with longer delay of KR and no reminder. The same might not be true were the delay period of longer duration, i.e., a longer intertrial interval is required in order to increase the KR delay interval, and so Exp. 2 was designed.

EXPERIMENT 2

The design was similar to that of Exp. 1, except that the delay of KR and the intertrial interval were longer.

Method

Apparatus.—In the previous experiment it was necessary for *S* to make a number of separate and distinct turns of the knob before receiving KR. It was thought desirable to change the task to one which required but one relatively simple act. The Manual Lever was selected, a device which also made it possible to use a 3–5 sec. interval as the minimum delay of KR. Hereafter, this 3–5 sec. minimum is designated as I (immediate) KR.

The Manual Lever has been completely described elsewhere (**3**). Essentially, it is a vertically suspended 14-in. lever offering a resistance of 20 lb. The task was learning to displace the lever by 34°. Upon a signal, *S*

pulled the lever briskly and released it at the desired amplitude. The displacement was read and recorded by *E*, who then reported KR to *S*. The KR was the algebraic difference between the goal amplitude and the measure of response amplitude.

Procedure.—The timing of the major events is given in Table 2. The first term of the notation designates the delay of KR, the second term the intertrial interval. Of the three groups with an intertrial interval of 1 hr., one group had both immediate KR and a reminder KR about 15 sec. before the signal to pull again, another had immediate KR with no reminder, and the third group had KR delayed by 1 hr. Group I-20, differing from the others in that the intertrial interval was 20 sec., was included to ascertain the direction of the effect of intertrial interval.

Three trials were administered, the last two of which can be considered as influenced by KR. Though the number of trials was small, it was thought a sufficient number for greatly reducing the size of the initial error. Further, it is reasonable to suppose that any effect of KR delay would more likely appear early than late in practice, particularly when the intertrial interval is relatively short.

The order of administering the groups was by simple counterbalancing of the three 1-hr. groups, interpolating from time to time one or two *S*s from Group I-20. The *S*s required to pull once per hour were returned to special isolation rooms during the post-KR period. Upon returning for the second or third pull, these *S*s were quizzed as to their general understanding of the task to be learned.

Results

The initial mean absolute error was about 12° of arc, and the error for Trial 3 was close to 7° of arc. As in Exp. 1, the Trial 2 data were used in an analysis of variance, and again the *F* for the between-groups variance failed to reach the .05 level of significance. Numerically speaking, Group I–20 was best, and the differences between the three other groups were small.

It was concluded that it would be difficult to demonstrate differential effects of the experimental variables with variation within an hour, but that another attempt should be made through manipulation of the post-KR period.

TABLE 2

TIMING OF WITHIN-TRIAL EVENTS FOR EXP. 2

Group	*N*	KR Delay	Reminder	Intertrial Interval
I–20 sec.	55	3 sec.	—	20 sec.
I–1 hr.	40	3 sec.	—	1 hr.
I–1 hr. (R)	40	3 sec.	1 hr.*	1 hr.
1 hr.–1 hr.	40	1 hr.*	—	1 hr.

* Administered about 15 sec. before signal to respond again.

EXPERIMENT 3

Experiment 3 was designed to investigate the effects of variation in duration of both post-KR and intertrial interval. The technique used held delay of KR constant at 5 sec. and made use of reminder KR. The values of the variables ranged up to 1 hr. as did those of Exp. 2. Experiment 3, however, is a bit more powerful than its predecessor in that there is an additional value of the experimental variable which might be more useful in detecting a trend. Furthermore, a greater number of *S*s per group improved the reliability of the data.

Method

Procedure.—The apparatus was the same as that used in Exp. 2. The details of the procedure were patterned after those employed in Exp. 2. The timing of the within-trial events is shown in Table 3. Inspection of this table will show that the first three groups vary intertrial interval from 20 sec. to 2 min. and 1 hr. The difference between the last three groups in Table 3 lies in the post-KR reminder period. This period takes the values 15 sec., 5 min., or 60 min.

Results

Statistical analysis of the data failed to reveal significant differences between groups. On Trial 3 the

TABLE 3

TIMING OF WITHIN-TRIAL EVENTS FOR EXP. 3 (KR is immediate for all groups and $N = 60$)

Group	Intertrial Interval	Post-Reminder Interval
20 sec. –15 sec. (R)	20 sec.*	15 sec.
2 min.–15 sec. (R)	2 min.*	15 sec.
1 hr. –15 sec. (R)	1 hr.*	15 sec.**
1 hr. –5 min. (R)	1 hr.	5 min.**
1 hr. –1 hr. (R)	1 hr.	1 hr.**

* Variable intertrial interval; delay of KR and post-reminder interval held constant.
** Variable post-reminder interval; delay of KR and intertrial interval held constant.

response error increases progressively with increases in duration of intertrial interval, but on Trial 2 there is a reversal. As the post-KR duration is increased, the Trial 2 means decrease progressively, but on Trial 3 there is a reversal. Consistent with the results of the previous study, Group 20 is numerically superior to any other.

On the basis of the results of Exp. 2 and 3 there is no evidence of important effects of variation in KR delay and post-KR delay within periods up to 1 hr. The data suggest that with a determined effort positive results can be obtained by varying intertrial interval.

EXPERIMENT 4

In this experiment the intertrial interval and delay of KR were increased to 24 hr. (one pull per day). Unless otherwise indicated, all points of procedure were the same as in Exp. 2 and 3.

Method

Three groups, all with a 24-hr. intertrial interval, were defined: Group I–24 hr., KR was immediate; Group I–24 hr. (R), KR was immediate and a reminder (R) of the previous score was given about 15 sec. before the signal to pull again; Group 24 hr.-24 hr., KR was delayed 24 hr., i.e., administered 15 sec. before the signal to pull again. Fortunately, it was possible to extend the number of trials to five for all *Ss*. The design is summarized in Table 4.

Results

Visual inspection of the response errors over successive trials suggested that learning had been impaired by the greatly increased value of the intertrial interval. The learning or *trials effect* within each group was evaluated by means of three separate subject-by-trials arrangements of the error data The trials effect, evaluated with the *Ss* × Trials mean square, gave two significant values of F (at the .01 level). The single insignificant value (F less than unity) was for Group I–24 hr., the group with immediate KR and no reminder. Thus, superior performance was obtained from a group with a long KR delay and a small post-KR delay as opposed to a group with immediate KR and long post-KR delay.

Because the statistical test for learning within Group I-24 hr. was negative, the data of each *S* were carefully inspected and compared for an internal check. A comparison, for example, of Trial 1 vs. Trial 5 errors showed only 12 out of 23 *Ss* with a lesser Trial 5 error, a finding in accord with the nonsignificant value of F. Trial-by-trial variability was positively correlated with the means, yet the largest ratio of variances was small (approximately 2).

TABLE 4

TIMING OF WITHIN-TRIAL EVENTS FOR EXP. 4

Group	*N*	KR Delay	Reminder Delay	Intertrial Interval
I–24 hr.	23	3 sec.	—	24 hr.
I–24 hr. (R)	24	3 sec.	24 hr.	24 hr.
24 hr.–24 hr.	23	24 hr.	—	24 hr.

Despite the insignificant value of *F* obtained, it is not necessary to conclude that the conditions represented by Group I–24 hr. precluded learning in all *Ss*.

By using the means for the immediate KR groups of Exp. 2 and 4, a crude evaluation of the retardation effected by increasing intertrial interval was made. As this interval increases from 20 sec. and 1 hr. to 24 hr., mean errors of 5.6, 6.5, and 8.2, respectively, are obtained. In the foregoing comparisons, there is an increasing post-KR interval with increasing values of the intertrial interval. Only in Exp. 4, however, was there evidence of positive effects attributable to the post-KR period.

Experiment 5

Experiment 4 gave the first evidence of a positive effect of the temporal position of KR. The outcome was most directly opposed to the view that KR delay is the most potent of the time variables, and favored the view that the post-KR interval is a relevant variable when the intertrial interval is long, i.e., 24 hr. Because the outcome of Exp. 4 differs, not only from orthodox expectation but also from earlier findings, it may reflect sampling error only, or an interaction of trial-duration and post-KR delay. While the writers had hoped for results similar to those of Exp. 4, they do not press the generalization. Of the alternative reliability checks, repeating the 24-hr. interval, and increasing the interval still further, the latter was chosen. As learning might be seriously impaired by further widening of the intertrial interval, the demonstration was considered risky. In the absence of learning there could be no differences between various delay groups. But this would at least establish the significance of the intertrial interval.

In Exp. 5 the intertrial interval was increased to 7 days to permit a comparison between immediate KR plus a 7-day post-KR interval and a 7-day delay of KR plus a comparatively insignificant post-KR interval. Unfortunately, it was impossible to continue the program with the Manual Lever. The task ultimately selected, learning to pull a yardstick 10 in. from a sheath while blindfolded, appeared to possess the essential characteristics of the lever task.

Method

Apparatus.—A wooden sheath was fastened to a laboratory table, and an ordinary yardstick inserted 35 in. into the sheath. The remaining inch overhung the table top and served as a grip for thumb and forefinger. While seated, *S* reached for the grip and, at the "pull" signal, partly removed the yardstick from its sheath. The right hand was used, the arm was fully extended, and the movement from left to right was fairly smooth and brisk. The *E* recorded the response to the nearest $\frac{1}{16}$-in. deviation from the 10-in. goal. Thus, for example, a pull of 9 in. was reported as 16 *units* short.

Procedure.—The details of procedure were patterned after those previously described. Only two treatments were designed, Groups I and D. Group I was given immediate KR, and reappeared for the next response 7 days later. Group D waited 7 days for KR. Under these rather severe scheduling conditions, it was possible to administer a total of 4 trials with *N*'s of 29 and 27 for Groups I and D, respectively.

The *Ss* did not remove the goggles until 15 sec. after response completion. Upon returning for another trial, *Ss* were reminded of the general nature of the task before them. After this briefing, the goggles were fitted. Fifteen seconds after the goggles were fitted, *Ss* of Group I began the 3-sec. foreperiod. At the goggles-fitted point, *Ss* of Group D were given KR for the response of the previous week, and 15 sec. later entered the foreperiod. The 15-sec. period served both groups as a brief adaptation period during which whatever cues *Ss* use in such experiments could be reinstated. This period also defines the post-KR duration for Group D.

Results

Inspection of the data for Trial 1 showed that the groups were not comparable in central tendency and variance. A careful review of the experimental design offered no rational reason for the observed discrepancy. The Trial 1 data were given to a second party who was able to compose 20 matched pairs of *Ss*. Only the data of the paired *Ss* are discussed below.

The mean absolute errors and their *SD*'s are to be found in Table 5. The effects of the two treatments are not easily summarized. Probably the key statement is that the variabilities are large. Though most *Ss* appear to improve, some do not, while a few get worse. As shown in Table 5, the trend in the means is downward and suggests an over-all learning effect. There appears to be no differential effect of treatment, in that Group D has the lesser mean on Trial 3, and Group I has the lesser mean on Trial 4. These differences are not significant.

Analyses of variance failed to reveal any significant Trials effect in four trials or in the last three. Apparently, a 7-day post-KR treatment is equivalent to a treatment of 7 days delay of KR. Though the locus of KR within the 7-day period has no differential effect, the 7-day period as represented in the intertrial interval does seem to retard learning. The lack of significance for the Trials effect supports the view that there was no learning, while the trend in the means suggests that there was. Certainly, the learning effect is not strong, and is generally weaker than that obtained in the earlier studies with lesser intertrial intervals. It is ventured that given four trials and a 7-day intertrial interval, the effect of the latter is to produce relatively large individual differences manifest in differing trends in the data of individual *Ss*.

TABLE 5

MEAN ABSOLUTE ERRORS IN $\frac{1}{16}$ IN. AND *SD*'s IN EXP. 5

Trial	Group I		Group D	
	Mean	*SD*	Mean	*SD*
1	59.9	40.5	60.0	43.4
2	56.6	53.4	61.0	40.5
3	43.8	40.0	42.8	48.0
4	31.4	28.3	39.2	35.2

Note.—Groups are 20 matched pairs of *Ss*.

DISCUSSION

Very few published studies on the effects of temporal delay of KR exist, especially if studies interpolating some activity between response and KR are excluded. There is reason to believe that many more studies have been run than published, an observation that suggests positive findings have proved difficult to obtain.

Of the two published experiments yielding positive results, that of Greenspoon and Foreman (6) is the more convincing. On the basis of eight *Ss* in each of four delay groups, they reported significant differences among delays of 0, 10, 20, and 30 sec., and slower rates of learning with increasing delay. The task was 3-in. line drawing with *S* maintaining an unsupported and tiring grip of the pencil during the pre- and post-KR periods. The authors point out that "hand-maintaining activity" may reduce the effectiveness of KR in that the KR would tend to "reinforce" the hand-maintaining activity as opposed to the act of drawing the line. If these data are reliable, the pre-KR activity places the experiment in a category of studies with filled (identical, similar, or different responses) delay intervals. In such relatively complex S-R-KR systems there are known marked deteriorations of response proficiency with lags in feedback. The foremost examples are

transmission lag in a continuous tracking task (16) and the trials-delay technique reported by Lorge and Thorndike (8) and by I. Bilodeau (4).

The present writers believe that while returning the hand to rest is certainly disrupting, the total amount of interfering activity is less than results from hand-maintaining activity. The difference in amount of interfering activity interpolated between response and KR's of different delays is also presumed less for hand-resting than hand-maintaining treatments. Because of the above considerations, the impracticality of the hand-maintaining technique, and the desirability of keeping the procedure as close as possible to that of animal reward, hand-maintaining was not used here.

The other positive finding is that of Saltzman (13) who concluded that a 6-sec. delay in information about the correctness of a choice response in a verbal maze retarded learning, but pointed out that his procedure allowed for delayed information about incorrect responses as well as more rehearsal of incorrect responses for the delayed group. Moreover, in Saltzman's design the longer delay group also had the longer intertrial interval. Macpherson, Dees, and Grindley (9) reported deterioration of performance with increasing values of intertrial interval—a finding in accord with the data of the present series of studies.

Restricting the discussion to reasonably controlled experiments with KR-delay periods not deliberately filled with interfering activities, there remain only the published studies of Lorge and Thorndike (8), Saltzman, Kanfer, and Greenspoon (14), and the unpublished studies of Alexander (1) and Noble (10), all of whom found no positive effect of delay in KR. On a counting basis, at least, the weight of the evidence supports the view that delay of KR is not itself an effective variable. In this context, Noble's studies deserve particular note as complementing the unusually long delays of the present series. Our shortest delay was of the order of 3–5 sec., because *E* required this time to read, round, and report *S*'s score. Noble carefully investigated the effect of delay with several values between 0 and 3 sec. in a task involving learning to push in correct order either 4 or 10 buttons. No differences attributable to delay were discovered.

The present findings show that a simple response can be related to its KR over much longer time intervals than hitherto believed. If delay of KR did matter, it was overridden by more powerful effects. Because the KR-delay interval did not seem to matter and because the value of the intertrial interval did, we have the question of what can be forgotten in such experiments. The possibilities, to be taken singly and together, are (*a*) the previous response, (*b*) the previous KR, and (*c*) the planned response correction (12). None of these seems critical when the intertrial interval is relatively brief. With wide variation in intertrial interval, forgetting of the previous response would seem to be a relevant factor. With long intertrial intervals and relatively long post-KR delays, the planned response correction may suffer.

Delay of KR vs. delay of reward.— More recent investigations of temporal delay of reward with animal *S*s have revealed gradients, in that the longer the delay the more the learning is retarded (7, 11). The notion of a KR gradient has been uncritically expected in the human motor skills context on the assumption that delay of KR is analogous to the delay of reward (food, water) in the animal studies. The few studies on delay of KR which are methodologically closest (none is identical) to those on delay of reward, however, have not produced gradients, but null results similar to the earlier reward studies.

Indeed, while investigators of human behavior have been attempting to make but a simple demonstration that delay of KR is an effective learning variable, the gradient argument from the rat laboratory has undergone an important change. Spence (15), for example, proposed that there is no gradient of "pri-

mary" reinforcement, and that the phenomena of delayed-reward learning be considered as effects of secondary reinforcement. An important point seems to be that it is futile to seek *the* shape and *the* slope of the gradient of delay of reward (goal gradient). Rather, shape and slope are functions of several variable factors, some temporal and some not. For example, no gradient at all may be expected when reward is delayed and secondary reinforcement is immediate.

The point that workers with rats should at will be able to produce almost any kind of gradient seems to have escaped many KR writers, the latter giving the impression that a matching gradient can be produced. So far, we have failed to see deduced from the animal work the conditions under which a KR-delay gradient can be found. And it should be acknowledged that there is nowhere a plot of such a gradient, unless studies with concomitant variation of other variables are admitted.

Summary

Five experiments were reported where the variables were the durations of delay of KR, post-KR delay, and intertrial interval. The salient features were the use of durations ranging from a few seconds up to seven days and of experimental designs with unconfounded temporal variables. Reminders of KR were introduced as a special control over the effects of forgetting, but the delay periods were otherwise unfilled.

Largely because there was no formal interpolated activity before and after KR and *S* learned a fairly simple response, little or no differential effect of delay of KR was expected. The results showed only null effects with delays of KR up to week-long intervals.

The evidence indicated that the critical variable is the intertrial interval, for performance varied inversely with its duration. Forgetting and, perhaps, a need to warm up account for the relationship.

The data suggest that the value of post-KR delay has more influence than KR delay. A group with immediate KR failed to learn with a 24-hr. intertrial interval, whereas learning was apparent in two other groups, one with immediate KR and a reminder 24 hr. later, and the other with KR withheld for 24 hr. However, with the intertrial period at seven days, within- and between-*S* variability become very large and it was impossible to demonstrate, in the few available trials, that learning had occurred in either delayed or nondelayed groups.

References

1. Alexander, L. T. Knowledge of results and the temporal gradient of reinforcement. Unpublished Ph.D. dissertation, Ohio State Univer., 1950.
2. Ammons, R. B. Effects of knowledge of performance: A survey and tentative theoretical formulation. *J. gen. Psychol.*, 1956, **54,** 279–299.
3. Bilodeau, E. A., & Ferguson, T. G. A device for presenting knowledge of results as a variable function of the magnitude of the response. *Amer. J. Psychol.*, 1953, **66,** 483–487.
4. Bilodeau, I. McD. Accuracy of a simple positioning response with variation in the number of trials by which knowledge of results is delayed. *Amer. J. Psychol.*, 1956, **69,** 434–437.
5. Brown, J. S. A proposed program of research on psychological feedback (knowledge of results) in the performance of psychomotor tasks. In *Research Planning Conference on Perceptual and Motor Skills.* San Antonio, Texas: Hum. Resour. Res. Cent., Lackland A. F. B., 1949. (Conf. Rep. 49-2.) Pp. 81–87.
6. Greenspoon, J., & Foreman, S. Effect of delay of knowledge of results on learning a motor task. *J. exp. Psychol.*, 1956, **51,** 226–228.
7. Grice, G. R. The relation of secondary reinforcement to delayed reward in visual discrimination learning. *J. exp. Psychol.*, 1948, **38,** 1–16.
8. Lorge, I., & Thorndike, E. L. The influence of delay in the after-effect of a connection. *J. exp. Psychol.*, 1935, **18,** 186–194.
9. Macpherson, S. J., Dees, V., & Grindley, G. C. The effect of knowledge of results on learning and performance: III. The influence of the time interval between trials. *Quart. J. exp. Psychol.*, 1949, **1,** 167–174.
10. Noble, C. E. Human delayed reward learning under two conditions of task length. Paper read at the Southeastern Psychol. Ass., Nashville, March, 1957.

11. Perin, C. T. A quantitative investigation of the delay-of-reinforcement gradient. *J. exp. Psychol.*, 1943, **32,** 37–51.
12. Reed, H. B. An experiment on the law of effect in learning the maze by humans. *J. educ. Psychol.*, 1935, **26,** 695–700.
13. Saltzman, I. J. Delay of reward and human verbal learning. *J. exp. Psychol.*, 1951, **41,** 437–439.
14. Saltzman, I. J., Kanfer, F. H., & Greenspoon, J. Delay of reward and human motor learning. *Psychol. Rep.*, 1955, **5,** 139–142.
15. Spence, K. W. The role of secondary reinforcement in delayed reward learning. *Psychol. Rev.*, 1947, **54,** 1–8.
16. Warrick, M. J. Effect of transmission-type control lags on tracking accuracy. Dayton, Ohio: Wright-Patterson Air Force Base, 1949. (USAF Tech. Rep. No. 5916.)
17. Wolfle, D. Training. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951.

(Received August 5, 1957)

Journal of Experimental Psychology
Vol. 55, No. 6, 1958

TEMPORAL EFFECTS OF CONDITIONED FEAR ON THE EYELID REFLEX [1]

K. W. SPENCE AND W. N. RUNQUIST

State University of Iowa

In a series of experimental studies (**4, 5**) evidence has been presented to show that the level of performance in eyelid conditioning of groups of *Ss* whose drive levels (*D*) were equated, nevertheless, varied with the intensity of the UCS employed on the conditioning (i.e., reinforcing) trials. This finding has been interpreted as supporting a reinforcement-type interpretation that habit strength, *H*, is a function of the intensity of the reinforcing UCS.

Drive level was equated in these studies by employing two different intensities of the UCS (air puff), each being present on half of the trials. In each of the first two experiments (**4**), two groups were compared under conditions in which the high reinforcement group received the strong UCS on the conditioning, i.e., paired CS-UCS, trials and the weak UCS on nonconditioning trials, i.e., either nonpaired trials or a temporal interval of pairing known not to produce conditioning. The reverse relations held for the low reinforcement groups. In the third experiment (**5**) all *Ss* were conditioned to two different CS's (light and tone). One of the CS's was paired with a strong UCS and one with a weak UCS, the particular CS being controlled.

While the level of *D* active at the time of occurrence of the anticipatory CR as determined by the UCS was thus controlled in these various experiments, the question may be raised as to whether differential *conditioned* emotional (fear) responses might have produced different levels of *D* in the high and low reinforcement groups or conditions. As Brown, Kalish, and Farber (**1**) have shown, conditioned anticipatory emotional responses presumably do become established in such aversive types of learning situations and do have motivational concomitants. Thus they demonstrated that a CS paired with a noxious UCS served to energize or intensify a reflex response made to another, different stimulus which followed the presentation of the CS on test trials by 3 sec.

Presumably in the present experiment the CS likewise becomes conditioned to an emotional response elicited by the noxious UCS. Moreover, it would seem reasonable to assume that this CR would be stronger in the cases in which the CS was paired with a strong puff than with a weak puff. In this event it is apparent that a stronger conditioned fear response and hence higher level of *D* might be operative in the case of the high reinforcement conditions than in the low reinforcement ones. Consideration was given to this possibility in the first article in this series (**4**), but it was suggested that the latency of such a conditioned emotional response would be much too long to effect the conditioned eyelid response, which occurs in our experiments from 200 to 500 msec. following the onset of the CS.

The present experiment was conducted for the purpose of checking this possibility that differential conditioned emotional responses might lead to different levels of *D* at the time of the occurrence of the CR and hence might be the factor responsible

[1] This study was carried out as part of a project concerned with the role of motivation in learning under Contract N9 onr-93802, Project NR 154–107 between the State University of Iowa and the Office of Naval Research.

for the differences in performance level that have been obtained. The design of the experiment was similar to that used by Brown and his colleagues (**1, 2, 3**). In the experimental group an emotional response produced by a shock UCS to the finger was conditioned to a light CS During the course of this conditioning a probe stimulus consisting of a weak puff to the eye was substituted for the shock on test trials spaced at intervals throughout the conditioning period. On half of these test trials the probe stimulus followed the onset of the CS by 500 msec. and on the other half by 4500 msec. A similar procedure was employed with a control group in which fear conditioning was precluded by using a backward order of presentation of the shock and the light. Our primary interest was in the question as to whether or not the experimental (fear-conditioned) group would exhibit an augmented reflex response to the probe stimulus at both the 500-msec. and 4500-msec. intervals as compared with the control (nonfear-conditioned) group. According to the reasoning given in the previous paragraph it would be expected that an augmented reflex response to the probe stimulus would be obtained with the 4500-msec. interval but not at 500 msec

Method

Subjects.—A total of 64 male students from an introductory psychology course served as *Ss*. They were assigned to one of two groups in order of appearance at the laboratory. The two groups differed only in the conditioning procedure. The experimental group had the CS onset before and overlapping the UCS, a brief shock; while the control group had the UCS before the onset of the CS, a procedure which presumably should not produce any conditioning of the emotional response to the CS.

Apparatus.—The *S* was seated in a semi-soundproof darkened room. The CS to which the emotional response was conditioned was an increase in the brightness of a 6-cm. circular disc from a level of .004 to .506 apparent ft.-candles. The duration of the CS on conditioning trials was 550 msec., with the shock UCS occurring 500 msec. after the onset of the CS for the experimental group and 500 msec. before the onset of the CS for the control group. The shock was provided by the discharge of a 4 microfarad condenser and was administered through electrodes attached to the index and middle fingers of the right hand. Amount of shock was regulated by means of a variable ratio transformer.

The response recorded was the eyeblink in the right eye to a weak air puff (.33 lb.) delivered to the left eye. The equipment for recording the eyeblinks and presenting the air puff was identical with that used in previous studies of this series (**4**). The duration of the air puff was limited to 50 msec. by means of a 110-v., 60-cycle AC-operated solenoid valve controlled by an electronic timer.

Procedure.—Following the reading of the instructions, the level of shock was adjusted for each *S*. The initial presentation of the shock was 50 v. and it was finally adjusted to either 80 v., 100 v., or 120 v., depending upon when *S* reported it to be unpleasant but not painful. All *Ss* were given 20 presentations of the air puff (PS) alone in order to adapt the eyelid response as much as possible. The experimental group was then given 40 light-shock pairings, and the control group 40 trials in which the shock preceded the light. The shock was increased 10 v. every 10 trials to prevent adaptation. Interspersed with the shock trials were test trials in which shock was omitted and the onset of the light was followed by the air puff. These test trials occurred following Trials 7 and 10 in each block of 10 conditioning trials, with the air puff following the CS onset by 500 msec. on one of these trials and by 4500 msec. on the other. On the latter trials the CS duration was 4550 msec. Two orders were used so that for each pair of test trials half of the *Ss* in each group had the short interval first and half the *Ss* had the long interval first. The order for half of the *Ss* was LS, SL, LS, LS. For the other half the order was, SL, LS, SL, SL. The intertrial interval on all trials was prearranged randomly at 30, 35, and 40 sec. A verbal ready signal preceded the onset of the CS by 1.5 to 3.5 sec. The *Ss* were instructed to blink and fixate the circular disc upon presentation of the ready signal.

Results

The experimental design provided a measure of the amplitude of the eyeblink to the air puff (PS) for each

S at two time intervals for each block of 10 conditioning trials. The mean amplitude in millimeters is plotted for both the 500-msec. and 4500-msec. intervals in Fig. 1. The zero point on the abscissa represents the mean of the median amplitude on the last three adaptation trials to the air puff alone. Actually, this measure is at best only a rough estimate of the preconditioning level of response since the CS was not presented on these trials. A Mann-Whitney test between the group means at this zero point failed to indicate any critical differences between the groups ($z = .09$).

The results of the fear-conditioning procedure clearly support the hypothesis that the latency of the conditioned emotional response is longer than 500 msec. While there is no evidence that the conditioned fear response had any effect when the PS was administered 500 msec. after the CS onset, the experimental group does show an increasing reflex facilitation over the control group when the PS occurred 4500 msec. following the onset of the CS. Both groups show an initially larger amplitude of response to the PS at the longer intervals; however, this difference has largely disappeared for the control group by the end of the experiment.

Since the data were markedly skewed, statistical evaluation of these results was made by means of a series of Mann-Whitney tests. The analysis of the effects of conditioning was made on the summed amplitudes for the last two PS trials at each interval. The combined amplitude for the short CS-PS interval was then subtracted from the combined amplitude for the long CS-PS interval. The mean difference was significantly greater for the experimental group than for the control group ($z = 3.05$, $P = .001$). The difference between the experimental group and the control group amplitudes at the 500-msec. interval was not significant ($z = .47$, $P = .32$), while the difference between these groups at the 4500-msec. interval was significant ($z = 1.87$, $P = .03$). The conclusion seems justified that an augmented reflex response to the PS was obtained with the 4500-msec. interval, but not at 500 msec.

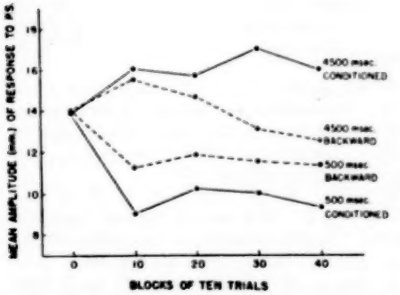

FIG. 1. Curves showing the differential facilitating effect of a conditioned emotional response upon the magnitude of eyelid closure to a probe stimulus (weak air puff) as a function of the time interval between the CS and the probe stimulus. The values plotted at the zero point on the abscissa represent the last three adaptation trials to the air puff alone.

DISCUSSION

It will be observed that the facilitating effects of the conditioned emotional response are largely manifested in preventing adaptation of the eyeblink response rather than by actually increasing blink size. Although an attempt was made to adapt this response to a low level before conditioning was started, there is evidence of additional adaptation in that the 500-msec. interval produced consistently smaller eyeblinks than those at the end of the initial adaptation period. The control group at the 4500-msec. interval appears also to drop after an initial rise while the experimental group maintains a fairly high level of response throughout. It should be pointed out that this initial rise and

drop of the control group has been obtained in another experiment using the probe stimulus technique for measuring the facilitation of an unconditioned GSR in human *Ss* (**3**). In this study an emotional response was conditioned by means of shock to light using a very long (20 sec.) CS-UCS interval. The GSR was measured to a brief clicking sound introduced after the onset of the CS. The control group in this experiment showed the same initial increase in the amplitude of response to the PS that was obtained in the present study, as well as the decrease in the amplitude of this response with further trials.

The findings of this experiment thus provide additional support for the assumption that a conditioned emotional response can be established to a neutral stimulus by pairing a noxious UCS with that stimulus. Consistent with the previous studies of Brown and his colleagues (**1, 2, 3**) the data also show that such a conditioned fear stimulus may act to energize or facilitate a reflex response. Furthermore it has been demonstrated that these energizing effects do not occur within 500 msec. after the onset of the fear CS within the conditions used in this study. If the results of this experiment may be generalized to the eyelid-conditioning situation, they support the assumption that the conditioned emotional response which develops to the CS in classical eyelid-conditioning experiments has too long a latency to influence the drive level at the time of occurrence of the eyelid CR. This finding adds further support to the interpretation that the differential performance of the high and low reinforcement groups in the present series of experiments was not due to different drive levels (*D*), but instead that habit strength (*H*) in classical aversive conditioning is a function of the intensity of the reinforcing UCS.

Summary

An experiment was conducted to determine whether a conditioned emotional response could augment a reflex response to another stimulus within 500 msec. in order to check on the possibility that differential conditioned emotional responses might lead to different levels of drive (*D*) at the time of occurrence of the CR in classical eyelid conditioning.

An emotional response produced by a shock UCS to the finger was conditioned to a light CS with a 500-msec. CS-UCS interval. During the course of conditioning, a probe stimulus (PS) consisting of a weak air puff to the eye was substituted for the shock on test trials spaced at intervals throughout the conditioning procedure. On half of the test trials the PS followed the CS by 500 msec. and on the other half of the test trials by 4500 msec. A control condition was employed which differed from the above only in that during conditioning the UCS preceded the CS by 500 msec., a procedure which is known to produce little if any conditioning.

The results showed that the experimental (conditioned) group exhibited a larger mean amplitude of response to the PS than the control group when the 4500-msec. interval was used but not when the 500-msec. interval was used. These findings were interpreted as supporting a position which assumes that habit strength in classical aversive conditioning is a function of the intensity of the UCS.

References

1. Brown, J. S., Kalish, H. I., & Farber, I. E. Conditioned fear as revealed by magnitude of startle response to an auditory stimulus. *J. exp. Psychol.*, 1951, **41**, 317–328.
2. Kalish, H. I. Strength of fear as a function of the number of acquisition and extinction trials. *J. exp. Psychol.*, 1954, **47**, 1–9.
3. Meryman, J. J. The magnitude of an unconditioned GSR as a function of fear conditioned at a long CS-UCS interval. Unpublished Ph.D. dissertation, State Univer. of Iowa, 1953.
4. Spence, K. W., Haggard, D. F., & Ross, L. E. UCS intensity and the associative (habit) strength of the eyelid CR. *J. exp. Psychol.*, 1958, **55**, 404–411.
5. Spence, K. W., Haggard, D. F., & Ross, L. E. Level of eyelid conditioning within the same subject as a function of different UCS intensities. *Psychol. Rep.*, in press.

(Received August 5, 1957)

Journal of Experimental Psychology
Vol. 55, No. 6, 1958

BINAURAL SUPPLEMENTATION OF THE AUDITORY INTENSIVE THRESHOLD

WILLIAM A. SHAW

University of Pennsylvania

For two ears of equal sensitivity, the probability of hearing a magnitude corresponding to the monaural threshold is increased by the product of the probabilities of hearing that magnitude with either ear alone. The magnitude of this effect (bias) Smith and Licklider (**4**) called "binaural supplementation." In discussing the possibility that the difference between the binaural and the equated monaural thresholds might be explained as the result of statistical bias rather than of summation, Smith and Licklider set up as a model the distribution formed by recording the better of "a pair of values drawn independently and at random" from a given normal distribution. The location of mean of the former with respect to the mean of the latter is known.

The location of the "binaural" mean by the use of the model is dependent upon the dispersion of the distribution from which the values are drawn. Unfortunately, reports in the literature up to the present show that the dispersion of differences between the binaural and the equated monaural thresholds are too small. Smith and Licklider believe that this is due to the method used in obtaining the data. If *S* is allowed ample time and, further, if he is permitted to search back and forth when near the threshold, each determination will be a kind of average. Over a number of such determinations the range will be smaller than it would be if *S* were not permitted such leisure. The dispersion of differences may also be too small because some kind of averaging over the duration of the stimulus takes place.

The purpose of the present study was to obtain a set of data which met the assumptions required by the model and which was free from the averaging thought to be inherent in existing data. For comparison another set of data was obtained from a single *S* for whom, although the assumption of independence could not be met, the method of collecting the data was again free from the more obvious kind of averaging. Lastly, a third set of data was collected for the purpose of studying the effect of duration on the dispersion of threshold measures.

METHOD

Part I.—The intensity of a 500-cps test frequency was divided into two electronically independent channels. A signal of 300-msec. duration was turned on and off by manually triggering an electronic switch adjusted to eliminate clicks. Both channels operated simultaneously and in phase into PDR-8 earphones equipped with circumaural cushions. Two 2 db per step attenuators mechanically coupled permitted simultaneous adjustment of the intensity of the signal in both channels. By means of a supplementary 1 db per step attenuator in each line the voltage across each earphone could be adjusted independently. Once set, the oscillator dial was not changed throughout the experiment.

All measurements were made with *S*s seated in soundtreated booths located in quiet rooms on each side of the experimental room. Occasional airplane noise could be heard by *S*s. During the time when such noise was present the experiment was stopped.

A signaling system was set up by means of which *E* could communicate with *S*s and they with him but not with each other.

The method of limits was used to determine the threshold. All test tones were preceded by a ready signal. Beginning with an intensity ranging from 6 to 12 db above threshold, *E* reduced this intensity in 2-db steps until neither *S* any longer signaled that a tone was heard. The *E* then adjusted the stimulus intensity from 6 to 12 db below threshold and increased it in 2-db steps until both *S*s reported hearing the tone. The exact starting point was varied for each new trial.

The threshold value for each *S* was recorded separately and was located halfway between the first change on the descending run and the first change on the ascending run. This value, minus from 6 to 12 db, was set into the attenuators for the next trial.

Preliminary training was given both *S*s. It consisted of eight sittings, each consisting of seven complete trials, i.e., seven descending and seven ascending runs. The sittings took place over a period of three weeks. The first trial was a practice trial and was not used. The next trial was used to establish the threshold referent for Trial 3. Trial 3 was used as the referent for Trial 4 and so on until the sitting was finished. By the end of the eight sittings, the results for both *S*s appeared to be stable and the experimental sittings were begun. Of these, there were 26, all identical to the preliminary sittings. During the experiment, *S*s were kept in ignorance of both the nature and the purpose of the experiment.

The *S*s were two graduate students with no previous experience in doing psychophysical work. Several threshold determinations were obtained for each ear of both *S*s. The better ear of each was used for further testing.

Part II.—The apparatus and procedure for the experiment on the single *S* was identical to that used in Part I except for the following: the mechanically coupled 2 db per step attenuators were driven by a motor controlled from a telephone-type switch located in *S*'s booth. A continuous 500-cps signal was used. The beginning intensity was adjusted by *E* as before. The *S* could not reverse direction during a run.

Twenty sittings of six trials each took place over a period of three weeks. A trial consisted of a descending and an ascending run. The monaural thresholds were determined first and these values were set into the attenuators for the binaural trial.

In Part II, as in Part I, once the trials were started no data were excluded except those of the practice trial at the beginning of each sitting.

The author served as *S*.

Part III.—Two frequencies, 500 cps and 6000 cps, were selected and each presented to *S* for durations of 100 and 800 msec. The stimulus conditions were presented in blocks of five trials each consisting of an up and a down run. The order of the stimulus conditions was such that each preceded and followed the other an equal number of times. The experiment took place over a period of six weeks following an intensive three-weeks practice period. The method of presentation was the same as that for the double *S* in Part I. During practice, *S*'s right ear was found to be the most sensitive. This ear was used throughout the experiment proper.

The *S* was a graduate student who had no previous experience in doing psychophysical work.

Results and Discussion

The results were treated in the following way: for each *S* a distribution of differences between successive thresholds within sittings was formed. The correlation between successive right–left pairs of differences was found to be very small (see Table 1). Since less than 4% of the variance in the two distributions of the double *S* is related, it was assumed that the two distributions were, for present purposes, independent. Therefore, the two distributions were combined.

TABLE 1

Means and *SD*'s (in Decibels) of Differences from Threshold

Cond.	*N*	Left Monaural		Right Monaural		Combined		*r*
		Mean	*SD*	Mean	*SD*	Mean	*SD*	
Double *S*	130	.20	2.22	−.12	2.24	.02	2.23	.19
Single *S*	80	.16	2.88	.00	2.21	.08	2.55	.16

TABLE 2

MEANS AND *SD*'S (IN DECIBLES) OF BEST OF PAIRS OF DIFFERENCES DRAWN AT "RANDOM" FROM THE COMBINED DISTRIBUTION OF SUCCESSIVE MONAURAL DIFFERENCES

Cond.	Mean		*SD*	
	Predicted	Obtained	Predicted	Obtained
Double *S*	1.24	1.11	1.85	1.99
Single *S*	1.43	1.49	2.12	2.12

In Table 1 are shown the means and *SD*'s of these distributions for each of the two *S*s of the double *S* in Part I and for the two ears of the single *S* in Part II. The *SD* for the left ear of the single *S* is inflated owing largely to the variability of a single final sitting which took place during the time *S* was recovering from a cold. In spite of this, all of the distributions were normal.

The successive right–left pairs were then treated as though they were drawn at random from the combined distribution of differences. The better of each pair was thrown into a distribution and the mean and *SD* found. These obtained values may be compared to those predicted by the model using Hastings, Mosteller, Tukey, and Winsor's constants (**1**). These authors show that the predicted *SD* should be .826 times that for the combined distribution given in Table 1. The predicted mean should be .564 *SD* units better than that for the combined distribution. The results for the double and single *S* are presented in Table 2. The agreement is reasonably good.

It is clear, however, that the values in Table 2 are not large enough to account for the 3-db difference which Smith and Licklider (**4**) cite as a typical difference between the monaural and binaural thresholds. The difficulty is amplified if it should turn out that the estimate of 3 db is too small. There is evidence in the article by Shaw, Newman, and Hirsh (**3**) that for 500 cps this is the case (viz., 5.6 db).

A new determination of the binaural gain was made using the same *S* who acted as the single *S* in Part II of this experiment. Fifty determinations of the binaural threshold were made. The mean difference between the equated monaural threshold and the binaural threshold was 4.72 db with an *SD* of 1.70 db.

Supplementation is not an explanation of the increased sensitivity shown by the binaural threshold unless it can be argued that the methods used still result in some kind of averaging, or that *S*s have some external cue which gives away *E*'s initial setting on the attenuators. Both of these would reduce the *SD* of differences. The rooms were well separated and there were no audible transients produced by the operation of the attenuators, hence neither of these suggestions is very likely, especially in view of the low correlation between pairs of right–left differences.

Smith and Licklider support their contention that some kind of averaging must be taking place by citing Lifshitz's (**2**) findings of a 6-db variation in the threshold within 5 sec. Apart from the difficulty of obtaining a valid measure of the threshold in 5 sec., deviations as large and even larger than 6 db did occur between successive determinations of the threshold for both the single and double *S*. It is obvious that differences of such magnitude do not occur frequently, and that 95% of the cases fall within 4.5 db. Moreover, when *S*'s determinations vary this much, he is aware of it and usually reports that the sitting was not very satisfactory, i.e., that he could not pay attention to the task or that he kept forgetting what the tone sounded like, etc.

It is possible, however, that the length of the signal might be a factor affecting

the amount of variability in the threshold. With regard to this variable, the *SD*'s for the double *S* who was presented with a 300-msec. stimulus are not very different from those of the single *S* who listened to a continuous stimulus. Since this finding could be attributed to differences in *S*s, a separate experiment (Part III) was conducted in which stimuli of different durations were presented to an *S* who had not previously been used.

For 500 cps the *SD*'s of successive differences for durations of 100 msec. and 800 msec. are 1.58 db and 1.84 db, respectively. The results for 6000 cps are similar. The *SD* for 100 msec. is 1.50 db and for 800 msec. it is 2.27 db. The variability tends to increase with the longer duration of the stimulus. Had the reverse of this been true, as Smith and Licklider seem to expect (**4**), there would be reason to suspect some kind of integration over the duration of the signal as a possible explanation of *S*'s apparent precision.

Further, the method used in this experiment to determine the absolute threshold did not allow *S* to search back and forth near the threshold as did the method used by Shaw, Newman, and Hirsh. Yet the results are closely similar. The change in method made little difference in the outcome. However, it made it possible to show that the binaural threshold cannot be generated from the independent monaural thresholds. There is reasonable certainty, therefore, that the increase in sensitivity shown by the binaural threshold over the equated monaural threshold is real and not merely apparent.

This increased sensitivity may be explained as the summation of the loudness produced by the simultaneous stimulation of the two ears of the single *S*. One would like to know, of course, the kind of neural events that take place in the auditory pathways leading to the cortex. This problem can be approached with greater confidence once there is assurance that the increase in binaural sensitivity is a real phenomenon.

Summary

Absolute monaural thresholds were obtained by the use of a modification of the method of limits with discrete stimuli from two *S*s simultaneously. These were compared to another set of monaural thresholds obtained by using the same method but with continuous stimuli to each ear of a single *S* successively. The results are highly similar. The binaural gain was greater than the increase in "sensitivity" obtained by choosing the better of the pair of independent monaural thresholds. This was true for both the double and single *S*. The amount of variability in the monaural thresholds is too small to be able to predict the increase in the binaural threshold over the monaural threshold by the use of the statistical model. The present experiment has shown that this is not an artifact of methodology nor of integration over the stimulus duration. The proposal of summation as an adequate explanation of binaural sensitivity is still highly plausible.

References

1. Hastings, C., Mosteller, F., Tukey, J. W., & Winsor, C. P. Low moments for small samples: A comparative study of order statistics. *Ann. math. Statist.*, 1947, **18,** 413–426.
2. Lifshitz, S. Fluctuations of the hearing threshold. *J. acoust. Soc. Amer.*, 1939, **11,** 118–121.
3. Shaw, W. A., Newman, E. B., & Hirsh, I. J. The difference between monaural and binaural thresholds. *J. exp. Psychol.*, 1947, **37,** 229–242.
4. Smith, M. H., & Licklider, J. C. R. Statistical bias in comparisons of monaural and binaural thresholds: Binaural summation of binaural supplementation. *Psychol. Rev.*, 1949, **46,** 278–284.

(Received August 12, 1957)

Journal of Experimental Psychology
Vol. 55, No. 6, 1958

PROGRESSIVE AMBIGUITY IN THE ATTAINMENT OF CONCEPTS ON THE WISCONSIN CARD SORTING TEST [1]

ISIDORE GORMEZANO AND DAVID A. GRANT

University of Wisconsin

The purpose of this investigation was to determine the difficulty of sorting for color and number in the Wisconsin Card Sorting Test (WCST) (**5, 6, 7**) when the irrelevant dimensions of number and form, respectively, were intermittently reinforced 0%, 25%, 50%, or 75% of the time. The study was also concerned with determining the transfer effects on number sorting of having intermittently reinforced the irrelevant number dimension during color sorting.

In concept formation *S* is generally confronted with a variety of stimuli objects each of which includes relevant and irrelevant elements. Relevant elements are consistently reinforced, but no such constant relationship holds for irrelevant elements. Some studies have attempted to determine the effects on concept formation of systematically varying the number of irrelevant elements (**1, 12**). Others have investigated the relative effectiveness of positive (relevant) and negative (irrelevant) instances on concept formation (**8, 13**), and some work has been done to assess the effects of the degree of intermittent reinforcement of the irrelevant elements which usually occurs in these experiments (**3, 4, 9**). These studies of intermittent reinforcement have concentrated on its possible role in the superiority of reversal shifts (where the relevant dimension stays the same, but the positive and negative signs are reversed) over nonreversal shifts (where a formerly irrelevant dimension becomes relevant). Buss (**3**) suggested that intermittent reinforcement of the formerly correct dimension impeded nonreversal shifts, but Kendler and D'Amato (**9**) and Buss (**4**) found that elimination of the intermittent reinforcement did not eliminate the *superiority of the reversal shift*. Their experiments were not designed to eliminate the possibility that intermittent reinforcement might enhance the *difficulty of the nonreversal shift* just as Buss conjectured. The present study provides evidence on this point.

Kendler and D'Amato (**9**) and Buss (**4**) interpret their findings as virtually requiring a mediational S-R theory (**11,** pp. 392–412) which assumes that in conceptual behavior *S* makes an implicit response which serves as the cue for the overt conceptual response. This theory, although it introduces some equivocation, supposedly permits an S-R continuity framework (**11,** pp. 446–455) to account for apparent noncontinuity where abstract behavior (**5, 6**) or verbal behavior (**10**) is involved. If intermittent reinforcement of an incorrect or irrelevant dimension during the initial stage of the present experiment leads to positive transfer on a second stage, where the formerly irrelevant dimension becomes relevant, the need for a mediational theory would be lessened as the

[1] This research was supported in part by the Research Committee of the Graduate School of the University of Wisconsin with funds provided by the Wisconsin Alumni Research Foundation.

results could be accounted for by conventional S-R theory. Absence of such positive transfer would tend to support the mediational position. The adequacy of the mediational theory, depending as it does on unobserved implicit responses which are postulated to behave in a rather peculiar manner, has not been tested. The theory hence remains an *ad hoc* and untested solution to the continuity—noncontinuity difficulty.

Procedure

General procedure.—The WCST, conventionally, involves the use of decks of 64 response cards (**5, 6**). In the present investigation, however, decks of 48 response cards of the "unsystematic" type (**7**) were used. Each card contained one to four figures of a single color. There were four colors—red, green, yellow, and blue; and four figures—triangles, stars, crosses, and circles. Each card could therefore be sorted according to the color, form, or number of the figures.

In the test situation a gray sorting tray with four double compartments was placed on a table before *S*. Four stimulus cards consisting of one red triangle, two green stars, three yellow crosses, and four blue circles were placed from left to right in the upper halves of the four double compartments. Two decks of response cards, differing in composition, were placed in a partitioned box fastened to one of the legs of the table, facing *E* and out of view of *S*.

The *S* was instructed in the following standard manner: "I am going to present you with a series of cards similar to those you see before you [*E* points to the stimulus cards]. Every time I give you a card I want you to place it in the compartment underneath the card to which you think it belongs. Each time you place a card I will tell you whether you are 'right' or 'wrong'." If *S* asked questions—i.e., whether to sort for color, number, or form, etc.—he was simply told: "I will tell you whether you are 'right' or 'wrong'." The *E* then selected a response card from one of the compartments and handed it to *S*. The response cards were always presented to *S* in a standardized manner and order. Initially the category color was called "correct," but after *S* had sorted 10 consecutive cards correctly according to color, *E* then began selecting response cards from the second deck, shifting the correct category to number. Care was taken to insure that *S* had no cue for the shift other than *E*'s statement of "right" or "wrong." After 10 correct number responses the experimental session was terminated. If *S* did not attain the criterion within 48 sortings in either stage of the experiment, his data were discarded.

Experimental design.—In the first stage *S* was assigned to one of four groups, each of which received one of four response card decks of 0%, 25%, 50%, or 75% number ambiguity; i.e., 0, 12, 24, or 36 of the 48 response cards, when sorted correctly by *S* on the basis of color and accordingly called "correct," would also have been "correct" on the basis of number. Each of the four initial groups was subdivided into four more groups in the second stage. Each of these four subgroups was given one of four response card decks of 0%, 25%, 50%, or 75% form ambiguity.

The decks of 64 response cards normally used in the WCST consist of all possible combinations of figures, colors, and numbers. In the present investigation, however, it was found that decks of 48 response cards (or multiples thereof) permitted the formation of eight decks of varying degrees of ambiguity that were balanced with regard to the various combinations of figures, colors, and numbers.

The experimental paradigm utilized in the present investigation is basically a complex variation of that of transfer of training, with a simple between-groups design of 4 groups, 40 *S*s in a group, in the first stage and a 4×4 factorial design, 10 *S*s in a cell, in the second stage. The virtues of the experimental design are several. First, the effect of degree of number ambiguity on color sorting in the first stage can be evaluated by a simple between-groups analysis of variance; second, the variation between columns of the 4×4 factorial design permits a determination of the effect of degree of form ambiguity on number sorting that is independent of any possible transfer effects of color sorting in the first stage; third, the between-rows variation provides a means of assessing the transfer effects to the second stage of the degree of number ambiguity during the first stage, which effects are again assessed independently of any transfer effects of color sorting.

Subjects.—The *S*s were 160 men and women in elementary psychology courses at the University of Wisconsin. Ten *S*s for each of the 16 cells satisfactorily completed the experiment. In addition there were 29 "failures" to complete either the first or second stage. These "failures" were distributed unsystematically over all experimental conditions, and the results obtained from these "failures" were discarded from the analysis.

Results

Each *S*'s performance in the two stages of sorting was recorded in terms of the following scores: (*a*) *total errors*; (*b*) *total trials*, i.e., total number of sortings required to attain the criterion, including the last 10 "criterion" cards; (*c*) the *total correct responses*, i.e., the number of correct sortings made by *S* before he began his criterion sequence of 10 successive correct responses; and, on the second stage only, (*d*) the number of *perseverative errors*, i.e., the number of incorrect responses which would have been correct in the first stage, except for the initial perseverative error which presumably served to inform *S* that a shift had taken place; and (*e*) the number of *nonperseverative errors*, i.e., the responses that also would have been wrong in the first stage of the experiment. By definition, there can be no perseverative errors in the first stage and nonperseverative errors simply become total errors.

The results of the experiment are summarized in Fig. 1, 2, and 3. In each of these figures the averages of all scores, after a square-root transformation, are plotted against the four levels of ambiguity.

Figure 1 presents the effect of increasing number ambiguity on color sorting. It can be seen from the average number of trials and correct responses that 25% number ambiguity had little effect on color sorting, but as number ambiguity increased to 50% and 75%, sorting for color became more difficult. The total number of correct responses required to attain the criterion indicates the number of reinforcements or reminders needed to confirm or establish the color response and has been considered a measure of the difficulty of sorting (**7**). The total errors appear to be unaffected by increasing number ambiguity. It is worth noting, however, that as number ambiguity is increased there is less opportunity for *S* to make errors, i.e., at 0% number ambiguity each time *S* sorts on the basis of

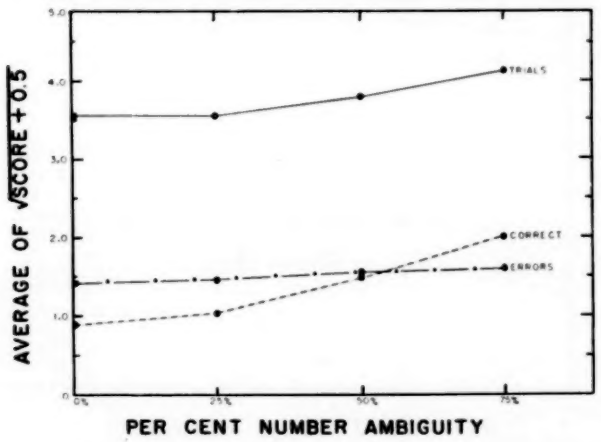

Fig. 1. The effect of increasing number ambiguity in color sorting upon total trials, total correct responses, and total errors.

TABLE 1

ANALYSES OF VARIANCE OF SCORES ON COLOR SORTING IN STAGE 1

Source of Variation	*df*	*F* Ratios		
		Total Trials	Total Correct	Total Errors
Progressive number ambiguity (between-groups)	3	10.9261*	13.5999*	.2423
Error	156	(.4075)	(.7353)	(.5237)

Note.—Numbers in parentheses are error mean squares.
* $P = .001$.

number an error is recorded, whereas at 75% ambiguity if *S* "erroneously" sorts all cards on the basis of number, only 25% of the responses could be recorded as errors.

Statistical analysis of the first stage of sorting was carried out by means of a simple analysis of variance, and that of the second stage by a rows-by-columns analysis of variance. Because all scores were frequency measures, it was considered desirable to use the transformation $g = \sqrt{x + .5}$ where g was the score analyzed and x the original score. This rational transformation serves to reduce the skewness of the frequency data, making them more amenable to the analysis of variance (**2**).

The summary of the simple analysis of variance of the three scores used to measure the effects of increasing number ambiguity on color sorting during the first stage is given in Table 1. Only the *F* ratios are presented, except for the error mean squares which appear in parentheses at the bottom of each column. The variation between groups was significant at the .001 level for total trials and correct responses, but the *F* for total errors was not significant.

Figure 2 depicts the effect of 0%, 25%, 50%, and 75% form ambiguity on number sorting in the second stage. Inspection of the figure indicates that all scores increased as form ambiguity went from 0% to 25%. However, only the total correct responses score demonstrates a uniformly progressive relationship to increasing form ambiguity. Thus, as form ambiguity increased from

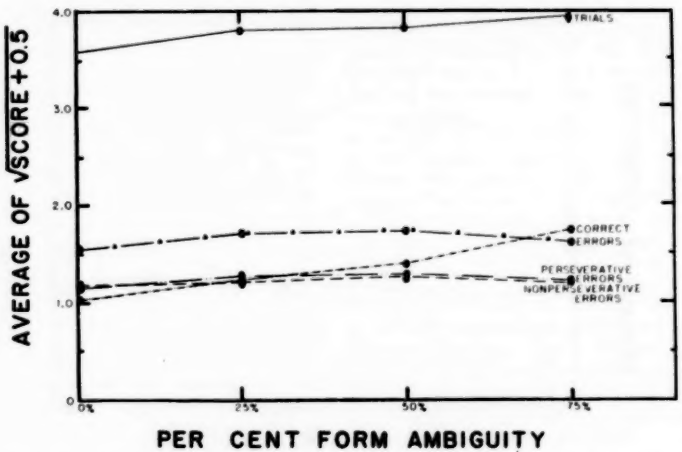

FIG. 2. The effect of increasing form ambiguity in number sorting upon total trials, total correct responses, total errors, perseverative errors and nonperseverative errors.

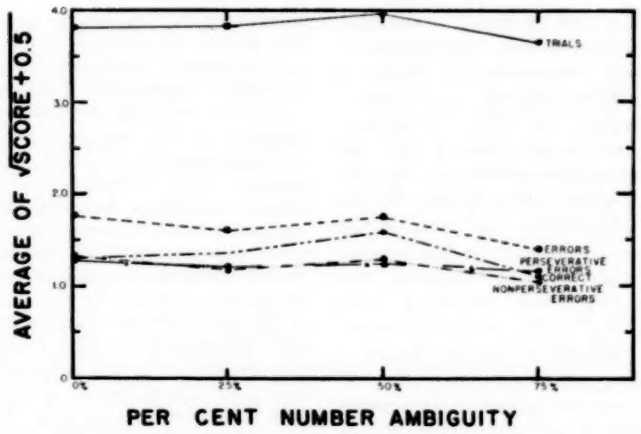

FIG. 3. The transfer effect of increasing number ambiguity in Stage 1 (color sorting), upon number sorting (Stage 2), for total trials, total correct responses, total errors, perseverative errors and nonperseverative errors.

0% to 75%, more correct responses or reinforcements were required for S to attain the criterion of 10 successive number sortings. The total trials, total errors, and perseverative errors showed no change as form ambiguity increased from 25% to 50%, whereas nonperseverative errors and total correct responses showed a rise. As form ambiguity went from 50% to 75%, total errors, perseverative errors, and nonperseverative errors decreased, but total trials and correct responses increased.

Figure 3 presents the effect of degree of intermittent reinforcement of number responses during Stage 1 (color sorting) upon subsequent number sorting (Stage 2). Except for total trials and total correct responses which show a slight rise, all scores used show a decrease as intermittent reinforcement of number rose from 0% to 25%. At 50% ambiguity all

TABLE 2

ANALYSES OF VARIANCE OF SCORES ON NUMBER SORTING IN STAGE 2

Source of Variation	*df*	*F* Ratios				
		Total Trials	Total Correct	Perseverative Errors	Non-perseverative Errors	Total Errors
1. Progressive number ambiguity in Stage 1 (row)	3	1.5407	1.8058	.2252	.6255	1.4066
2. Progressive form ambiguity in Stage 2 (columns)	3	1.3711	4.8072*	.4815	.1356	.5278
3. Progressive number and form ambiguity (rows × columns)	9	1.1536	.7546	1.6544	.4543	1.6333
4. Error	144	(.5826)	(.8157)	(.4334)	(.2836)	(.7524)

Note.—Number in parentheses are error mean squares.
* $P = .01$.

scores show an increase, indicating that 50% intermittent reinforcement of number responses in Stage 1 made number sorting in Stage 2 more difficult. At 75% ambiguity, however, all scores show a decrease below that of 0% number ambiguity.

Table 2 gives the summary of the analysis of variance for each of the five scores considered in the second stage of sorting. In Row 1, the source of variation tested is that shown in Fig. 2, the effect of degree of form ambiguity during Stage 2 on number sorting. Only the *F* for total correct responses is significant. Row 2 of this table shows that for all scores, the degree of intermittent reinforcement of number responses in Stage 1 of sorting had no statistically significant effect on number sorting in Stage 2 as was indicated in Fig. 3. In Row 3 the interactions between the Row 1 and Row 2 factors are tested. No significant interaction was found for any score.[2]

Discussion

The three principal findings of this study are: (*a*) increasing the degree of number ambiguity made sorting for color more difficult in Stage 1 of the experiment; (*b*) increasing the degree of form ambiguity made sorting for number more difficult in Stage 2; and (*c*) increasing the degree of number ambiguity and hence increasing the amount of intermittent reinforcement of the number responses on Stage 1, when *S*s were made to sort for color, had no significant transfer effect on number sorting in Stage 2.

The interpretation of the first two findings is straightforward. When the percentage of ambiguity is increased, this means an increase in the amount of intermittent reinforcement given the irrelevant dimension. The only basis *S* has for learning the correct sorting response is via the reinforcement or nonreinforcement of each of his responses. The relevant dimension is always reinforced, but with increasing ambiguity an irrelevant dimension is reinforced up to 75% of the time. Hence, increasing ambiguity makes it increasingly difficult for *S* to learn which of the two dimensions is correct. Buss (**3**) is therefore quite correct in assuming that intermittent reinforcement of an irrelevant dimension in a nonreversal shift would impede learning of the new response. Although intermittent reinforcement can hardly account for all of the superiority of reversal shifts over nonreversal shifts (**4, 9**), it can account for part of it.

The third finding was a negative one; it can become definitive only if the null hypothesis can be proved. Although this cannot be done, it is proper to point out that with 160 *S*s the power of the experiment is considerable, so that if intermittent reinforcement of an irrelevant concept affects future learning it must not affect it very much. This indicated failure of the overt responses to follow the continuity principle makes a mediational theory necessary to the S-R approach, but the S-R theorist must face the problem of noncontinuity in the implicit (verbal) mediating responses.

Summary

An investigation was made of the difficulty of sorting for color and number in the Wisconsin Card Sorting Test when the irrelevant dimensions of number and form, respectively, were intermittently reinforced 0%, 25%, 50%, or 75% of the time. The experimental design used also permitted a determination of the transfer effects on number sorting, of having intermittently reinforced the irrelevant number dimension during earlier color sorting. The design was basically a complex variation of that of transfer of training, with a simple between-groups design of 4 groups, 40 *S*s in a group, in Stage 1 and a 4 × 4 factorial design, 10 *S*s in a cell, in

[2] A second analysis of the data of Fig. 3 was carried out, using a covariance adjustment for total correct responses and total trials during the first stage to correct the second stage scores. This refinement led to no change in the significance levels of the *F*'s reported in Table 3.

Stage 2. The results and conclusions were as follows:

1. As number ambiguity was increased from 0% to 75% on Stage 1, sorting for color became more difficult.

2. Sorting for number became more difficult as form ambiguity was similarly increased on Stage 2.

3. Increasing number ambiguity or intermittent reinforcement of number-sorting responses in Stage 1 of sorting had no statistically significant transfer effect on number sorting in Stage 2.

4. It was concluded that intermittent reinforcement probably plays a supplementary role in the superiority of reversal shifts over nonreversal shifts in concept formation, and that the mediational S-R theory is supported although noncontinuity in the learning of the implicit mediating responses requires further explanation.

References

1. Archer, E. J., Bourne, L. E., & Brown, F. G. Concept identification as a function of irrelevant information and instruction. *J. exp. Psychol.*, 1955, **49**, 153–164.
2. Bartlett, M. S. The use of transformations. *Biometrics*, 1947, **3**, 39–52.
3. Buss, A. H. Rigidity as a function of reversal and nonreversal shifts in the learning of successive discriminations. *J. exp. Psychol.*, 1953, **45**, 75–81.
4. Buss, A. H. Reversal and nonreversal shifts in concept formation with partial reinforcement eliminated. *J. exp. Psychol.*, 1956, **52**, 162–166.
5. Grant, D. A., & Berg, E. A. A behavioral analysis of degree of reinforcement and ease of shifting to new responses in a Weigl-type card-sorting problem. *J. exp. Psychol.*, 1948, **38**, 404–411.
6. Grant, D. A., & Cost, J. R. Continuities and discontinuities in conceptual behavior in a card sorting problem. *J. gen. Psychol.*, 1954, **50**, 237–244.
7. Grant, D. A., & Curran, J. F. Relative difficulty of number, form, and color concepts of a Weigl-type problem using unsystematic number cards. *J. exp. Psychol.*, 1952, **43**, 408–413.
8. Hovland, C. I., & Weiss, W. Transmission of information concerning concepts through positive and negative instances. *J. exp. Psychol.*, 1953, **45**, 175–182.
9. Kendler, H. H., & D'Amato, M. F. A comparison of reversal shifts and nonreversal shifts in human concept formation behavior. *J. exp. Psychol.*, 1955, **49**, 165–174.
10. Kuenne, M. R. Experimental investigation of the relation of language to transposition behavior in young children. *J. exp. Psychol.*, 1946, **36**, 471–490.
11. Osgood, C. E. *Method and theory in experimental psychology*. New York: Oxford Univer. Press, 1953.
12. Reed, H. B. The learning and retention of concepts: IV. The influence of complexity of stimuli. *J. exp. Psychol.*, 1946, **36**, 252–261.
13. Smoke, K. L. Negative instances in concept learning. *J. exp. Psychol.*, 1933, **16**, 583–588.

(Received August 8, 1957)

Journal of Experimental Psychology
Vol. 55, No. 6, 1958

SEQUENTIAL DECISION MAKING: WALD'S MODEL AND ESTIMATES OF PARAMETERS[1]

GORDON M. BECKER

Electric Boat Division, General Dynamics Corporation

Human beings often make decisions with incomplete knowledge. Statistical test procedures to optimize decisions under uncertainty conditions (**2, 3, 6, 11**) have been adapted as models for describing the human decision process (**1, 3, 4, 5, 8, 9, 10**). The models usually force *S* to make his decision at a point predetermined by *E*.

A sequential probability ratio test devised by Wald (**2, 11**) permits the decision maker to secure as much information as he desires prior to making his decision. The procedure requires that the decision maker, prior to obtaining any sample data, specify (*a*) the specific finite set of mutually exclusive alternatives from which he will select his final decision, and (*b*) the α and β risks that will be associated with his final decision. The sample space can be described in two dimensions: the total number of items in the sample, and the number of sample items possessing a relevant discriminatory characteristic. This space is further divided into discrete decision areas. Associated with each decision area is the single decision or choice that will assure the decision maker of operating within his preselected risk levels. For two-choice problems, three decision areas are obtained: (*a*) accept the hypothesis, (*b*) reject the hypothesis, and (*c*) gather more information. These decision areas are separated by two straight parallel lines determined by the decision (population) characteristics and the preselected risk levels. If the sample is plotted in this space, item by item, as it is obtained, and if a decision to accept or reject the hypothesis is made as soon as the sample falls along either of the two criteria lines described above, the risk levels will not be greater than those originally specified and the sample will, on the average, be half as large as that required by nonsequential test methods.

This experiment was primarily designed to determine whether or not Wald's test could be used as a model for describing the human decision process. If this model is appropriate, then the decisions of an *S* permitted to obtain as much information as he desires prior to making a decision would fall along two straight parallel lines when plotted in Wald's space. Such an *S* would make decisions as soon as the α and β risks were each less than certain fixed limits.

In addition to the above test of Wald, previously unavailable data that might be helpful in deriving a sequential model of the human decision process were obtained. These included (*a*) the size of the sample at the time of decision, (*b*) the deviation of the sample mean from the parent population mean, (*c*) the probability

[1] This research was performed as part of the Basic Research Program of General Dynamics Corporation. The help and cooperation of A. E. Hickey, Jr., J. Vincent Harrington, T. G. Faria, and F. A. Brooks of Electric Boat Division are gratefully acknowledged. The author is greatly indebted to Lila Knapp for her assistance in selecting and supervising the statistical tests of Wald's model, and to Mortimer H. Applezweig and George O. Moeller for their cooperation in providing Connecticut College laboratory test facilities and students.

of a wrong decision, and (*d*) the probability of a correct decision. These measures were derived from theoretical sample distributions, not from the usual count of right and wrong responses.

A priori considerations suggested that the decision process might depend upon at least two problem characteristics: (*a*) the distance between the means of the parent populations, and (*b*) the number of parent populations from which *S* had to choose. These two characteristics were independently varied and analyzed in terms of the observed changes in the dependent measures.

Method

Apparatus

An electrical probability machine generated random sequences of items selected from one of seven dichotomous populations. Items from the selected population were generated one at a time by a two-way spring-centered switch in the center of *S*'s console. Generated items were recorded and accumulated on two 3-digit counters, one on each side of the switch. The counters were duplicated on the control console. The probability of the right (and left) counter firing was controlled by six toggle switches on *E*'s console.

Each of these toggles closed part of the circuit between a synchronous motor switch (Series RC, 6-switch synchronous motor driven, single cycle, multicam timer, Industrial Timer Corp.) and the left *or* right counter. Each motor switch closed once in $\frac{7}{8}$ sec., and remained closed for $\frac{1}{8}$ sec. Only one of the six switches was in the closed position at any instant of time. By properly choosing the number of toggles connected to a counter, the probability of that counter firing could be varied in seven equal steps from zero to one. The sum of the probabilities for the two counters always equaled one.

The switch on *S*'s console opened and closed the final circuits between the motor switches and the counters. Relays in these circuits immediately disconnected the counters from the circuit after either fired. Only one counter fired per activation of *S*'s switch depending upon which of the six motor switches was closed and *E*'s preset toggle connection. The motor switches opened and closed continuously.

Subjects

Eight paid volunteer undergraduate women from Connecticut College served as *S*s for eight weeks. The *S*s were tested two or three times a week depending upon their outside schedules.

Tests

The *S* drew as large a sample as needed to decide from which of two or more populations she was drawing. The possible populations from which the sample originated were described on a 3×9-in. white card located over *S*'s counters. A set of description cards was prepared for each problem series. Each card described two, three, four, or five populations, one of which generated the sample drawn by *S*.

Populations were randomly designated by letters A through F. In order to minimize letter preference effects, the letter designating a population varied from one presentation to the next. Each problem was presented eight times each session. The order of presentation, top vs. bottom card placement, and letter designations of the populations were randomized and equally probable within each test session.

Two different problems, varying in difficulty, were employed. The difficult problem involved a choice between a population containing 4000 left and 2000 right counter items and a population composed of 3000 left and 3000 right counter items. The easy problem consisted of a 3000/3000 and a 1000/5000 population. In addition, a third problem involving a 1000/5000 and a 6000/0 population was included as a filler item in the first problem series.

A second series of 24 two-choice problems involving different sets of population pairs was presented immediately after the above series. The population pairs comprising this series differed from *S* to *S* in a Latin-square pattern. Each *S* made a total of 48 decisions during each of the first 10 sessions.

A third series of problems, each involving more than two populations, was added to the program during Sessions 11, 12, and 13. This series consisted of four problems: a three-choice problem consisting of populations 1000/5000, 3000/3000, and 5000/1000; a three-choice problem consisting of populations 1000/5000, 3000/3000, and 4000/2000; a four-choice problem consisting of populations 1000/5000, 2000/4000, 3000/3000, and 4000/2000; and a five-choice problem consisting of populations 1000/5000, 2000/4000, .3000/3000, 4000/2000, and 5000/1000. The 48 problems of this series together with the 48 problems of the other series resulted in *S* making 96 decisions during each of the last three test sessions.

Procedure

On arrival, each *S* was shown her console, one population description card, and a printed set of instructions. The instructions told *S* how to operate the toggle switch to obtain sample items, explained the task of deciding from which of the described populations her sample originated, and informed her that she would be paid according to the combination of the number of right and wrong decisions she made and the total number of items she drew during the experiment. Every correct decision earned five points, every wrong decision deducted five points. The *S* with the highest ratio (total right-wrong score to the total number of items drawn) was paid $2.00 per test session and the *S* with the lowest ratio was paid $1.00 per session. Each series of problems scheduled for a session had to be completed during that session and all sessions had to be completed before *S* was paid. Each *S* was also informed that the problems differed from day to day and from *S* to *S*, but that these differences would be eliminated by a series of weights based upon the differences in the programs.

The *S* was not told her score or given any information about her choices until the end of the final test session. She was permitted to draw as many items as she wanted prior to each decision, but was not allowed to draw items after any decision. There was no way for *S* to directly learn her score or compare her performance with that of other *S*s.

After each decision, *S* returned the description card, reset her counters at zero and received a new card describing a different set of populations. The *E* reset the control console so that the next sample would be randomly generated from one of the populations described on the new card. This procedure was repeated without interruption until all problems and all series scheduled for the session had been completed.

Measures

The *S*'s decision and the number of sample items, by left, right, and total counter readings, were recorded for each problem. Three transformed scores were derived from these data: risk, hit, and criteria.

The risk score was derived from the cumulative binomial distribution (**12**) of the rejected population. The hit score was derived from the cumulative binomial distribution (**12**) of the accepted population. The criteria score was obtained from the theoretical *SD* and mean of the accepted population.

Risk.—The risk score estimated *S*'s α (or β) error, i.e., the probability that *S* made the wrong decision. This measure was obtained for every decision regardless of which population *E* actually presented *S*, and regardless of whether *S*'s decision was actually right or wrong. The transformation formula is given in Table 1. Two mean risk scores were determined for each problem, one for each decision. The individual score associated with a 3000/3000 decision on the difficult problem was the probability that a theoretical 4000/2000 sample of the size *S* actually obtained would have characteristics as deviant as those of the sample *S* decided came from a 3000/3000 population. On the easy problem, the risk associated with a 3000/3000 decision was the probability that a theoretical 1000/5000 population sample would have characteristics as deviant as those of the sample *S* called 3000/3000. The risks associated with the 4000/2000 and 1000/5000 decisions were similar theoretical probabilities that theoretical 3000/3000 population samples were called 4000/2000 and 1000/5000, respectively. Only two theoretical populations were considered for each decision, regardless of whether *S* actually selected from two, three, four, or five populations.

Hit.—Hit scores estimated the probability that *S* had made the correct decision, e.g., the hit score associated with the 3000/3000 decision was an estimate of the probability that a theoretical 3000/3000 sample of the same size would have characteristics as similar to the 3000/3000 population as those of the sample *S* decided came from the 3000/3000 population. The transformation formula is given in Table 1.

Criteria.—Criteria scores, defined in Table 1, estimated the deviation of the sample from the

TABLE 1

THEORETICAL MEASURES AND PARAMETERS

Measure	Formula
Risk (y_j)	$\sum_{i=d(y_j)}^{Z} \binom{S(y_j)}{i} p_X^{\,i}\, q_X^{\,S(y_j)-i}$
Hit (y_j)	$\sum_{i=d(y_j)}^{Z} \binom{S(y_j)}{i} p_Y^{\,i}\, q_Y^{\,S(y_j)-i}$
Criteria (y_j)	$\dfrac{d(y_j) - p_Y\, S(y_j)}{(S(y_j)\, p_Y\, q_Y)^{\frac{1}{2}}}$

Note.—p_X = the number of right (or left) counter items in population *X* divided by the total number of items (right + left counter) in population *X*. $q_X = 1 - p_X$. $S(y_j)$ = total number of items (right + left counter) in sample *j* that *S* decided came from population *Y*. Z = zero (0) when p_X is greater than p_Y; $Z = S(y_j)$ when p_X is less than p_Y.

mean of the parent population. Criteria scores of zero indicate S accepted a population when the sample mean was exactly equal to the population. Criteria scores of -3 indicate that S accepted a population when the sample mean was three SD below the population mean. In order to reduce the effects of sample size on the criteria estimates, the score was computed by dividing the actual distance between the sample and population means by the SD of the accepted population.

It is important to note that all three derived measures were based upon theoretical binomial distributions, rather than the usual count of right and wrong responses. It was possible with this procedure to estimate each measure after every decision rather than obtain, in the usual manner, only a single estimate from the entire set of decisions.

Results

Wald

The results for each S on the two, two-choice, Series *1* problems for 13 days were plotted in Wald's space. Figure 1 shows the results for two S's, S_6 who closely approximated Wald's model and S_1 who deviated most from Wald. Visual inspection of these graphs suggests a similarity to Wald, especially on the difficult problem. A summary of the statistical test results is given in Table 2.

Linearity.—Lines of best fit were computed for each of the eight Ss on

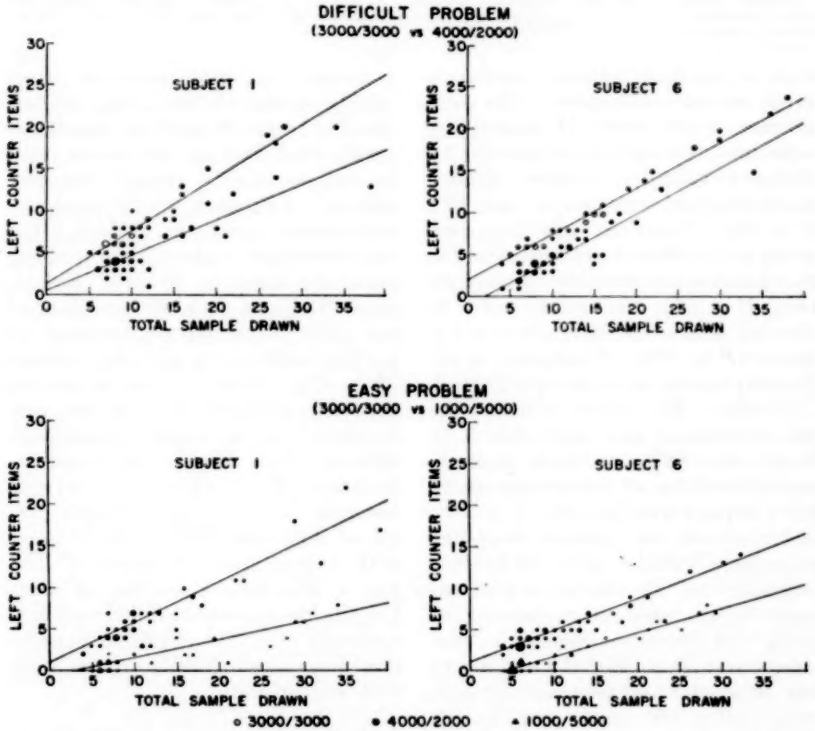

FIG. 1. Performance of two Ss plotted in Wald's decision space. S_1 deviated more from Wald's sequential sampling procedure than did any other S. S_6 was among those who most closely conformed to Wald.

TABLE 2

SUMMARY OF WALD'S TEST RESULTS

S	Difficult (Wald's Slope = .585)				Easy (Wald's Slope = .317)				Both
	df	3000/3000‡	4000/2000‡	//†	*df*	3000/3000‡	1000/5000‡	//†	All Slopes
1	40	**.423**	**.625	**	46	.482**	**.223**	**	**
2	48	.540	.534**		46	.433**	.442**		**
3	38	**.473**	.585	*	48	.407**	.325	*	
4	44	.498*	.668**	**	47	**.479**	**.199**	**	**
5	46	.498**	.655**	**	46	.414**	.254**	**	**
6	39	**.553	*.549		52	.375**	.293*	**	
7	52	.611	.565		47	**.384	**.360		
8	41	.563	.615		42	*.444*	.398		
Mean	8	.520	.600		8	.427**	.312**	**	**

Note.—Asterisks indicate statistical difference at .05(*) and .01(**) levels. Over-all test results shown in bottom row and last column are based on Kolmogorov-Smirnov one-sample test (7).

† Asterisks indicate statistical level of *F* test of parallel lines (6) based on differences between the two obtained slopes.

‡ Tabled numbers are the slopes of the lines best fitting *S*'s obtained performance. Asterisks preceding slope refer to significance of *F* test of linearity (6). Asterisks following slope refer to significance of *t* test (6) of difference between obtained slope and Wald's predicted slope.

each of the four decisions, and separately tested for linearity. An over-all test of the entire 32 lines led to rejection of straight-line behavior for the performance as a whole (Kolmogorov-Smirnov one-sample test [7], $P < .01$). However, this departure could be attributed predominately to S_1 (Kolmogorov-Smirnov one-sample test, $P < .01$), the other seven *S*s showing essentially straight-line behavior ($P > .05$). A summary of the linearity test results is given in Table 2.

Parallel.—The slope of each line was determined and compared, *S* by *S* (*F* test [6]), on each separate problem. Four of the difficult-problem line pairs were parallel ($P > .05$), and three of the easy-problem line pairs were parallel ($P > .05$). Performance on the difficult problem could be accepted as conforming to Wald (Kolmogorov-Smirnov one-sample test, $P > .05$); the easy problem ($P < .01$) and performance as a whole ($P < .01$) could not be. A summary of the statistical tests for parallel lines is given in Table 2.

Slope.—The slope of each line was compared to that predicted from Wald's model (*t* test, [6]) to test whether or not *S* used the population means described on the choice cards as well as Wald's general test procedure. A summary of the statistical test results is given in Table 2. The decision slopes on the difficult problem were the same as Wald's (Kolmogorov-Smirnov, $P > .05$) and four of the eight *S*s could be accepted as having Wald's slope on both problems ($P > .05$). However, the slopes on the easy problem and on the performance as a whole statistically differed from Wald (Kolmogorov-Smirnov, *P*'s $< .01$). The average absolute deviation from Wald's predicted slope was .073 for the 32 lines, with a maximum deviation of .165 and a minimum deviation of zero. Despite the statistical difference, it is apparent that the objective population means and Wald's process were both approximated.

Risk, Hit, Criteria, and Sample Size

Risk, hit, criteria, and sample-size measures were determined for every

TABLE 3

MEAN SCORES BY *Ss*

S	Measure			
	Risk	Hit	Criteria	Sample Size
1	.09	.57	.24	9.35
2	.04	.61	.21	17.56
3	.07	.54	.33	10.70
4	.04	.53	.35	13.38
5	.05	.59	.19	13.77
6	.07	.58	.29	10.13
7	.05	.49	.52	8.72
8	.09	.58	.30	9.38
Mean	.06	.56	.30	11.62

decision for each *S* during the last three test sessions. The mean scores for each *S* are shown in Table 3, and the mean scores for each condition are shown in Table 4. Kendall rank correlations between sample size and risk were computed for *Ss* as a group for each decision under each condition. All the correlations were small and negative. Neither correlation within the difficult problem was significant, but the easy problem 3000/3000 correlation (−.571) was significant beyond the .05 level. Individual Pearson product-moment correlations for each of the first three *Ss* within the easy problem were also determined. None of the correlations, within these *Ss*, was significant at the .05 level.

TABLE 4

MEAN SCORES BY CONDITIONS

Condition			Measure			
Alternatives	Prob.	Decision	Risk	Hit	Crit.	Sample Size
2	Diff.	4000/2000	.131	.495	.36	12.94
		3000/3000	.163	.513	.31	11.66
2	Easy	3000/3000	.029	.596	.19	10.86
		1000/5000	.060	.597	.27	9.83
3		1000/5000	.035	.581	.29	10.22
4		1000/5000	.021	.597	.33	12.48
5		1000/5000	.018	.541	.38	13.42

However, all three correlations were negative (−.57, −.07, and −.10) indicating that for a given problem there may be a weak tendency to reduce risk with increased sample size.

Risk.—Significant differences were found in the variance from condition to condition and it was therefore decided that nonparametric tests would be used throughout the remaining analysis. Despite equal objective costs and pay-offs, *Ss* uniformly had higher mean risks on the 1000/5000 decision than on the 3000/3000 decision within the easy problem (Wilcoxon matched-pairs signed-ranks test [7], $P < .05$). No consistent preference was shown within the difficult problem.

The mean risk on the difficult problem, 3000/3000 choice, was larger than the mean risk on the easy problem, 3000/3000 choice, for every *S*. This difference in risk as a function of problem difficulty was also reflected in the combined means of each problem, all *S*'s taking more risk on the more difficult problem (Wilcoxon matched-pairs signed ranks [7], $P < .01$).

Most *Ss* took more risk on the 1000/5000 choices when the number of alternatives decreased (Friedman analysis of variance by ranks [7], $P < .01$). However, it should not be inferred that risk absolutely decreased with an increase in the number of alternatives. The computed risks were estimates of the probability that *S* decided a theoretical 3000/3000 sample came from the 1000/5000 population. The *S* was more likely to have considered the 2000/4000 and 1000/5000 populations rather than the 3000/3000 vs. 1000/5000 population choices when there were five alternatives available, and *S* most likely considered this same possibility under the four-choice condition.

Significant differences in the mean risk at the time of decision were found between *Ss* (Friedman analysis of variance, $P < .01$). The ordering of *Ss* was the same regardless of the number of alternative choices in the problems they solved.

Hit.—All *Ss* had higher mean hit probabilities on the easy problem than they did on the difficult problem. The difference within the 3000/3000 choices was significant at the .05 level (Wilcoxon matched-pairs signed-rank test [7]). This increase in theoretical hit probability with increased distance between the means of the alternative populations suggests a relation between the usual concept of "difficulty" and the statistical concept of "*d*" (**10**).

None of the other independent variables results in consistent changes in the hit measure. The risk preference shown within the easy problem was not reflected, and *Ss* did not show a consistent rank order from one condition to another.

Criteria.—The variance of this measure was so large as to preclude significant differences either between the means of the different conditions or between *Ss*.

Sample size.—Most *Ss* took larger samples on the difficult problem than they did on the easy problem (Wilcoxon matched-pairs signed-ranks test, $P < .01$) but this difference was not consistent within the 3000/3000 decisions. The risk preference shown within the easy problem was not reflected in the sample-size measure.

Sample size increased with the number of alternatives given *S* (Friedman analysis of variance by ranks, $P < .05$). This was in agreement with the changes in the hit and risk measures.

Differences between *Ss* were significant within the two-choice problem and over-all conditions (Friedman analysis of variance by ranks, $P < .01$). The differences were more pronounced than those shown in either the risk or hit measures.

Discussion

The results show that people do not precisely follow Wald's model. However, despite the statistical difference, the model appears to be a reasonable first approximation: the slopes are close to those predicted, the lines are approximately parallel, and the linear correlations are large. These findings suggest that the solution to a given problem is based upon a constant likelihood ratio or betting odds. This consideration apparently determines the sample size desired prior to making a decision. Man's decision process also appears to depend upon two criteria, one for each decision, rather than a single criterion assumed in other decision-making models.

The obviously poorer fit to the data on the easy problem indicates that Wald's process is, at most, only a partial description of the human process. The changes in risk and hit probabilities from problem to problem and the decision choice preferences within each problem show that people do not simply maximize hits or minimize risks or maintain constant objective utility.

The decrease in risk and increase in hits with increased distance between the population means indicates that the criteria space moves with the population means. This suggestion is substantiated by the insignificant change in criteria from one problem to another, and within each separate problem. The apparent constancy of the criteria measure suggests that *S* sampled until the sample mean approximated one of the population means. If this were true, the slopes of the decision lines would have approximated the population slopes, i.e., the 3000/3000 decision line slopes would be close to .5, and the 4000/2000 decision line slopes would be close to .667, etc. The deviations of the ob-

tained slopes from the population slopes were therefore determined for each *S* on each decision to test this hypothesis. The average, absolute deviation was found to be .083 as compared to .073 from Wald's model. Within the difficult problem the averages were .057 and .056 for the population and Wald models, respectively, and within the easy problem the deviations were .109 and .090, respectively. The slope predictions from both models in respect to each other and to the obtained results were so close as to prohibit a crucial test.

However, data had been collected, in the second problem series, involving a problem in which *S* chose between a 5000/1000 and a 2000/4000 population. The slope predictions on this problem significantly differed from one model to the other. The actual deviations obtained from *S*s also differed significantly and favored Wald's model (Wilcoxon matched-pairs signed-ranks test, $P < .05$). The average, absolute deviations were .093 for Wald and .240 for the population model.

All *S*s drew at least four sample items before making a decision. However, if one guessed at the solution before drawing a single item, the denominator of the ratio used to compute the final score would have been zero. This would result in a score of either plus or minus infinity, and assure that *S* of receiving either the highest or lowest score. This strategy[2] would give a 50/50 chance of winning the $2.00/hr. payment. No *S*s adopted this apparently ideal approach. All preferred the one chance in eight that they could make better and faster decisions than the other seven *S*s. This implies that people do not rely on a pure objective game strategy or, at least, that *S*s used in this experiment overestimated their own abilities as compared to their estimates of the other contestants in the game.

[2] The author is indebted to W. P. Tanner, D. M. Green and J. A. Swets for these considerations.

Summary

Subjects were permitted to secure as large a sample as they desired prior to deciding from which of several populations their sample originated. Their performance was analyzed in terms of Wald's sequential probability ratio test model, in terms of three theoretical measures—α and β error levels, hit probabilities, and criteria—and in terms of the size of the samples drawn by *S* prior to a decision. It was concluded that:

1. Although Wald's model does not precisely predict performance, the discrepancies are of such small magnitude as to permit the use of Wald's model as a first approximation to the human decision process. Predictions are better on difficult than on easy problems.
2. Despite equal objective costs and pay-offs, and despite equal objective probabilities, *S*s show preferences for the type of error they make on easy problems. The pattern of these preferences are consistent from *S* to *S*.
3. The *S*s took less risk and were more likely to be right on easy problems than they were on difficult problems. They took less risk when there were more alternative choices, but their hit probability remained essentially constant.
4. The amount of risk and size of sample preferred at the time of decision differed from *S* to *S*. The relative ordering of *S*s in terms of these parameters was independent of the number of alternatives involved in the decision.
5. No *S* used the "ideal" game strategy.

References

1. Bilodeau, E. A. Statistical versus intuitive confidence. *Amer. J. Psychol.*, 1952, **65**, 271-277.
2. Dixon, W. J., & Massey, F. J., Jr. *Introduction to statistical analysis.* New York: McGraw-Hill, 1951.
3. Edwards, W. The theory of decision making. *Psychol. Bull.*, 1954, **51**, 380-417.
4. Edwards, W. Reward probability, amount, and information as determiners of sequential two-alternative decisions. *J. exp. Psychol.*, 1956, **52**, 177-187.
5. Irwin, F. W., & Smith, W. A. S. Value, cost, and information as determiners of decision. *J. exp. Psychol.*, 1957, **54**, 229-232.
6. Mood, A. M. *Introduction to the theory of statistics.* New York: McGraw-Hill, 1950.

7. SIEGEL, S. *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill, 1956.
8. SWETS, J. A., TANNER, W. P., JR., & BIRDSALL, T. G. The evidence for a decision-making theory of visual detection. Ann Arbor, Mich.: *Tech Rep. No. 40*, Electronic Defense Group, Univer. of Michigan, 1955.
9. TANNER, W. P., JR. A theory of recognition. Ann Arbor, Mich.: *Tech. Rep. No. 50*, Electronic Defense Group, Univer. of Michigan, 1955.
10. TANNER, W. P., JR., & SWETS, J. A. A decision-making theory of visual detection. *Psychol. Rev.*, 1954, **61**, 401–409.
11. WALD, A. Sequential tests of statistical hypotheses. *Ann. math. Statist.*, 1945, **16**, 117–186.
12. Tables of the cumulative binomial distribution. *Annals of the Computation Laboratory of Harvard University*, 1955, No. 35.

(Received for early publication
October 14, 1957)

Journal of Experimental Psychology
Vol. 55, No. 6, 1958

THE RELATIONSHIP BETWEEN DISCRIMINABILITY AND GENERALIZATION: A RE-EVALUATION [1]

HARRY I. KALISH

Adelphi College

Despite the relative absence of empirical studies, many investigators (**10, 12, 14**) have tacitly assumed a relationship between discriminability and stimulus generalization. For the most part, the nature of the relationship is considered to be inverse, i.e., generalization occurs to the extent that the organism fails to discriminate. Despite the rational tenability of these assumptions, however, recent studies by Guttman and Kalish (**7, 8**) which explored the process of generalization along a wave-length dimension utilizing the pecking response in pigeons, failed to confirm the inverse hypothesis and revealed an interesting paradox. The nature of the wave-length continuum, and the discriminability function ($\Delta\lambda = (f)\lambda$) derived from it, suggested that certain characteristics of this function might be reflected in the generalization gradient depending upon the wave-length value of the stimulus chosen as the CS. The general form of the discriminability function for pigeons (**9,** Fig. 1) suggests the specific predictions for the four stimuli (530, 550, 580, and 600 mμ) used as CS values. Where $\Delta\lambda$ is small (530 mμ), the gradient should be sharp; where $\Delta\lambda$ is large (530 mμ), the gradient should be flat. For 550 and 600 mμ where $\Delta\lambda$ is either increasing or decreasing, the gradients were predicted to be asymmetrical. In general, none of these predictions was realized. Instead, analysis of the empirical gradients (see Fig. 1) indicated that the form of the gradient

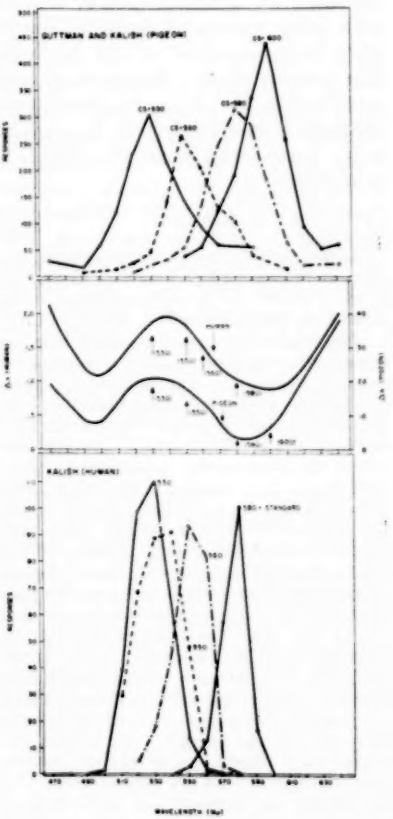

FIG. 1. Upper: Mean generalization gradients for pigeon (**7**). Middle: Hue discrimination as a function of wave length for pigeons (**9**) and humans (**5**). Lower: Generalization gradients of total responses for all Ss to the variable stimuli.

[1] This study was supported by grant M 1459 from the National Institute of Mental Health, United States Public Health Service. The author is indebted to Robert Berryman for his helpful suggestions and Russ Sardo for his assistance in the research.

was quite similar in each instance and very nearly symmetrical. These results implied that the underlying discriminability function for such gradients should be a line parallel to the abscissa rather than the prevailing sinusoidal-like function. Because this extraordinary lack of correspondence between discriminability and generalization process has important ramifications in the study of behavior it is necessary to conduct similar studies with other species in order to determine the generality of the findings. The present study, therefore, is concerned with the discriminability-generalization relationship in human *Ss* using wave length as the stimulus continuum.

Method

Subjects.—Forty female undergraduate students at Adelphi College were used as *Ss*. All *Ss* were tested for color blindness with the Farnsworth Dichotomous Test (**4**).

Apparatus.—The apparatus used in previous experiments with pigeons (**7**) was adapted for use with human *Ss*. The modified apparatus consisted of a plywood compartment painted flat black approximately 15 in. long, 11 in. high, and 14 in. wide. The *S* sat in front of the open side of the compartment approximately 33 in. away from a ¾-in. aperture in the center of the opposite wall. A plastic disc mounted behind the aperture served as a stimulus source and was illuminated by a 6-v., 18-amp. ribbon filament lamp whose beam was directed through a Cambridge Thermionic Corp. Monochromator which provided variations in wave length from 450 to 640 mμ. The characteristics of the monochromator are described in an earlier study (**7**). The *S* was provided with a telegraph key which closed a shutter placed between the monochromator and the projection lamp when the key was released. The shutter was automatically closed by a Hunter Decade Interval Timer if *S* did not respond within a 3-sec. period. An attempt was made to keep extraneous sources of light at a minimum by closing two sets of window shades, but no effort was made to achieve completely lightproof surroundings.

Procedure.—The procedure followed in the present study for presentation and judgment of the stimuli was a variation of the constant method known as the Method of Single Stimuli (**6**). Essentially, the method consists of *S*'s judgment of a variable stimulus without the standard stimulus for comparison.

After *S* was seated in front of the stimulus source the following instructions were read. "This is an experiment in color perception. At the beginning of the experiment a specific color will be presented through the small white hole in front of you. Try to keep this color in mind because you will be asked to identify it later. After 1 min. this color will be turned off and you will place your hand and press down on the telegraph key in front of you. I will give the signal 'ready' and a few seconds later a color will again be presented. You must decide whether this is the original color shown you at the start of the experiment. If it is, lift your hand as rapidly as you can from the key. If it is not, keep pressing on the key. I will say the word 'ready' whenever I am about to present a color and you should be pressing the key at that time. We are going to try some practice trials. Now we are going to run through a series exactly as we would do it if this were the real experiment. [*E* exposes 600 mμ for approximately 1 min. and then presents 610, 590, 600, 620, and 610 mμ.] Now we are going to begin the experiment. Remember, try to keep the original color in mind and respond as rapidly as you can, lifting your hand only when the original color appears. Do not be disturbed, however, if you should respond to other colors." Six randomized series of eight variable stimuli plus the standard were presented to *S* after he had viewed the standard stimulus for a period of 1 min. The total possible number of responses for each stimulus, therefore, was six and the total possible number for the six series of nine stimuli (eight plus the standard) was 54. Each variable stimulus was presented for a period of 3 sec. and was followed by a blackout of approximately 5 to 10 sec. allowing sufficient time for *E* to record the presence or absence of a response and change the monochromator setting. The 40 *Ss* were divided into two groups of 20 and each group was assigned to two of the four wave lengths chosen as standards. The *Ss* were selected for Group A or B successively as they appeared for the experiment. Group A was given 530 and 560 mμ and Group B, 500 and 580 mμ as standards. In each case the variable stimuli consisted of four wave lengths above and below the standard separated by intervals of 10 mμ. One half of Group A received 530 as the first standard and 560 as the second, while the other half of the group received the counterbalanced order. The same procedure was followed for Group B with 550 and 580 as standards. Both

test series for each group were conducted in one experimental session and approximately 10 min. rest was permitted between the test series.

Results

The combined graphs presented in Fig. 1 summarize the results from two studies. The gradients in the uppermost graph, showing the relationship between wave length and stimulus generalization for the pigeon, were obtained from the study by Guttman and Kalish (**7**). The discriminability functions immediately below this graph reflect the spectral difference thresholds for pigeons and humans and were used to generate predictions regarding the form of the generalization gradient. The remaining graph consists of the results obtained in the present experiment and represents decrement in total response for each stimulus relative to the particular series and standard employed. The *S*'s scores for each stimulus for the six test series were summed and the total for all *S*s plotted above the appropriate stimulus. Each of these curves may be construed as generalization gradients around the standard stimulus. Since the form of the discriminability function is appreciably the same for humans as it is for pigeons, similar predictions regarding the shape of the generalization gradients were made.

One of the expectations from $\Delta\lambda$ relative to the shape of the gradient is an increasing asymmetry resulting from changes in slope on the left side in the direction of decreasing wave lengths. Specifically, the slopes of the gradient should become progressively flatter until the 530 series at which point the gradient should approach symmetry with some tendency toward flatness in the direction of increasing wave lengths. The curves in Fig. 1 indicate that, to a very marked extent, these expectations were fulfilled. Using the standard stimulus as the point which divides the gradient in half, it is apparent that the number of responses for the left half of the gradients increases with decreasing wave lengths indicating that the slope of the gradient changes in the predicted direction. It is also obvious that, as discriminability decreases (the ascending portion of the discriminability function involving 560 and 550 mμ), judgment becomes more difficult and more confused resulting in a shift in the maximum number of responses from the standard stimulus to stimuli in the direction of decreasing wave lengths. The greatest number of responses for the 560 series occurs at 550 mμ, and similarly, maximum responding for the 550 series occurs at 540 mμ. The only exception to the observed correspondence between $\Delta\lambda$ and stimulus generalization is the 530 series where asymmetry should have been produced by flatness in the direction of increasing wave lengths, but where instead the opposite results were obtained.

Additional support for the reliability of these observations was obtained from trend tests of analysis of variance (**11**). All measures for each *S* were converted by subtracting the number of responses to each stimulus from the number of responses to the standard in order to achieve more uniform distributions. The F obtained for the gradients on the right side ($F = 12.20$, 9 and 228 df, $P < .001$) and left side ($F = 6.05$, 9 and 228 df, $P < .001$) suggested that these gradients differed with respect to slope. Individual analyses for pairs of gradients on the left side revealed that all comparisons were significantly different with respect to slope ($P < .001$) except for 530–

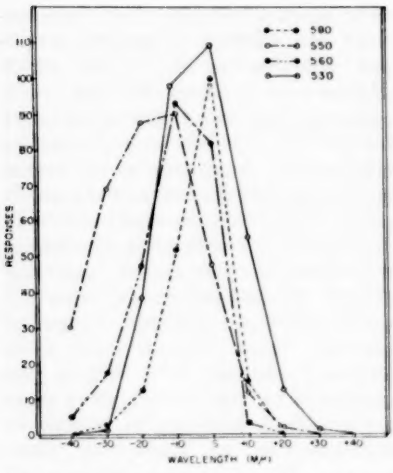

FIG. 2. Generalization gradients of total responses plotted around a common standard.

560 for which $F < 1.00$. For the gradients on the right side, only those comparisons with 530 attained significance ($P < .001$).

The second set of predictions from the discriminability function is concerned with over-all level of responding, i.e., the vertical displacement of the gradients which reflects the ease or difficulty of discrimination over the chosen series of wave lengths. The graph in Fig. 2 was constructed to emphasize differences in vertical displacement by plotting all of the gradients around a common standard. Differences in slope are also more readily apparent from this figure. Predictions from the $\Delta\lambda$ function relative to over-all level of responding imply that the gradients become broader as the wave length of the standards decrease. The two most extreme examples for the series in the present study occur at 530 and 580 mμ where, due to large differences in the absolute value of $\Delta\lambda$, the gradient around a standard of 530 should completely envelop the 580 gradient. Figure 2 discloses that, in general, these results were obtained. The reliability of these conclusions was demonstrated through the use of median chi square (**13**). This consisted of recording the total number of responses made by each *S* on the six trials and arranging these values in a distribution of the frequency of occurrence of total scores in the four groups as in Table 1. Median chi-square tests were performed by combining the two distributions to be tested, finding the median of the combined distribution, and arranging 2×2 chi-square tables with groups as one dichotomy and above and below the median as the other. The results of this analysis are recorded

TABLE 1

DISTRIBUTION OF FREQUENCY OF OCCURRENCE OF TOTAL SCORE ON SIX SERIES

(Possible Total Score = 54)

Total Scores on Six Test Series	Frequency of Occurrence of Scores in Each Group			
	530	550	560	580
4				1
5				
6			1	4
7				2
8		1		2
9	1		1	1
10	2		1	4
11		2	3	2
12	2	1	4	
13	1	2	3	2
14	2		3	1
15	1	1	2	
16	4	3		
17	2	1	1	1
18	1	3	1	
19		3		
20				
21		1		
22	2	1		
23				
24	1	1		
25				
26				
27	1			
Total	20	20	20	20

TABLE 2

MEDIAN CHI-SQUARE TEST VALUES FOR COMPARISONS OF OVER-ALL LEVELS OF RESPONDING

	Standard (mμ)		
	550	560	580
530	1.00	8.18	11.37
550		6.40	9.64
560			8.08

Note.—P = .02 for chi square of 5.41 for 1 *df*.

in Table 2 and show that, with the exception of the 530–550 comparison, all of the paired analyses validate the expectations from Δλ. Over-all level of responding appears to increase with decreases in wave length.

A more relevant analysis of the accuracy of the predictions emerging from Δλ is accomplished by deriving some function from the empirical data analogous to the discriminability function and comparing it with the Δλ curve. It was proposed[2] that the absolute value of Δλ for a given standard should conform to a measure of the band width of the gradient for any given response level for that standard. The values in Fig. 3 show the band width in millimicrons relative to the standard in millimicrons. The empirical curves were obtained by selecting the band width for the gradients at various response levels, determining the mean band width, and plotting these against the standard. The empirical values for the uppermost graph in Fig. 3 were obtained from Guttman and Kalish (7) and represent the mean band widths for the range of responses from 50 to 400 inclusive for the four standards. The values for human *Ss* from the present study represent the mean band widths for the range of responses from 10 to 80 inclusive. The values for the hypothetical curves were obtained from the discriminability function for pigeons and humans respectively (Fig. 1). They represent the absolute value of Δλ at the given standard stimulus. The curve for humans labeled "80 Responses" was included to show the remarkable degree of correspondence between this and the hypothetical function. Both the empirical and hypothetical values were adjusted to the same ordinate.

Despite the deviation introduced by the empirical value at 550 mμ for humans, the general shape of the empirical curve for mean band width appears to conform quite closely to the hypothetical curve. A test of goodness of fit by means of analysis of variance (11) yielded an $F < 1.00$ which substantiates the observed similarity of the two curves. The same test of goodness of fit was applied to

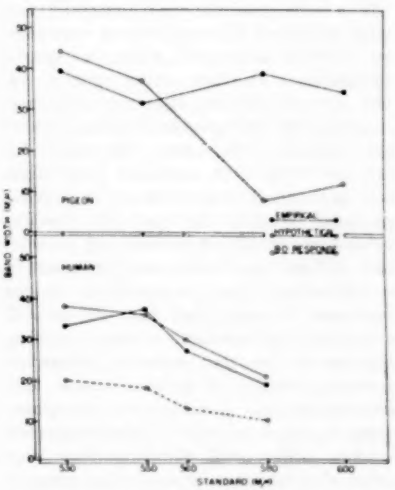

FIG. 3. Mean band width (empirical) and absolute value of Δλ (hypothetical) as a function of the standard for pigeons (upper) and humans (lower).

[2] The author wishes to thank Clarence Graham for suggesting this method of analysis.

the data for pigeons, and in this instance the null hypothesis could be rejected ($F = 5.52$, 3 and 19 *df*, $P < .01$).

Discussion

The generalization gradients obtained in the present experiment furnish striking evidence for the supposition that the processes of generalization and discrimination, as generally defined, bear an inverse relationship to each other and that they are fundamentally dependent upon the characteristics of the underlying stimulus continuum. The meaningfulness of the assertion that the present study demonstrates this relationship, however, depends upon the extent to which the method of single stimuli can be regarded as representative of the operation commonly associated with generalization in contrast to the methods used to obtain the discriminability functions.

The decision to use the method of single stimuli in this study grew out of the rationale involved in the experiment by Brown, Bilodeau, and Baron (3) where gradients of generalization were obtained from voluntary responses to stimuli arranged along a spatial dimension. In the same paper (3) the authors observe that generalization experiments and psychophysical studies are basically the same, the only difference being that empirical generalization is reduced to a minimum in psychophysical studies by use of training procedures which, it is assumed, parallel the differential reinforcement used in conditioning. An examination of the literature reveals that the use of differential reinforcement during training appears to be the essential difference between studies in generalization and discrimination. For example, the operations required to obtain discrimination involve differential reinforcement and either simultaneous or successive presentation of stimuli. In generalization experiments no differential reinforcement is involved; instead, training is directed toward raising *S*'s response strength to a given stimulus by varying the number of reinforcements. The procedure during testing for discrimination may or may not involve a comparison stimulus; in stimulus generalization a comparison stimulus is never present. The major difference between generalization and discrimination, therefore, appears to center around the administration of differential reinforcement during training.

The discriminability functions used in the present experiment for both the pigeon and the human were obtained by methods which achieved discrimination training through the use of either differential reinforcement or comparison stimuli. For the pigeon (9), training was designed to maximize the differences between two hues and threshold measurements were obtained by changing the wave length continuously until *S* responded similarly to both stimuli. In the case of the human $\Delta\lambda$ function (5), one side of a bisected field of a given wave length was gradually changed until a point was reached where the two wave lengths could be distinguished in hue.

The procedure for obtaining gradients of generalization for the pigeon (7) appears to be straightforward since customary conditioning techniques were involved. The basis for human generalization in the present study derives from the definition of empirical generalization prescribed by Brown, Bilodeau, and Baron as instances where "an organism that has been trained (or instructed) to respond to a designated stimulus will also respond, under certain specifiable conditions, to formerly neutral stimuli on which no training has been given (or to stimuli to which responses have been prohibited by instructions)" (3, p. 56). The *S*s in the present experiment were instructed to "attend" to a standard stimulus of a given wave length and were then required to select the standard from among a number of randomly presented stimuli of differing wave lengths. As in the case of Brown, Bilodeau, and Baron (3), decrements in the frequency of response to stimuli other than the standard were regarded

as instances of stimulus generalization. Apparently, then, the usual distinction made between the procedures designed to yield discrimination and those used to obtain generalization apply to the methods used in this study.

The supposition that generalization and discrimination are inverse processes is not new and this experiment has corroborated what other investigators have assumed (**10, 12, 14**) or demonstrated (**1**). What is essentially important, however, is that the discriminability-generalization relationship appears to be inconsistent with the data derived from the study with pigeons (**7**). This apparent discontinuity in the attributes underlying pigeon and human behavior is particularly interesting because it appears only under those conditions where responses are elicited to a single stimulus. The differences do not appear in the case of the discriminability function where both the human and pigeon *S*s are required to make a choice between two stimuli presented simultaneously. It is also interesting to observe that these similarities and differences occur despite differences in techniques used with the two species. Conditioning procedures were uniformly used to obtain both the discriminability function and the generalization gradients for pigeons (**7, 9**) while instructions were used with humans (**5**). This suggests that the differences in experimental method may not adequately account for the obtained differences in generalization since there is no appreciable difference in the form of the discriminability function where similar differences in method exist.

One of the several possibilities which may be raised to account for the differences in generalization results is that the discriminability function for the pigeon is not a correct representation of the spectral difference threshold. This possibility is highly unlikely, however, since Hamilton and Coleman (**9**) used standard conditioning procedures and obtained virtually the same results for three *S*s. Moreover, the similarity between the discriminability functions for humans and pigeons is not entirely unexpected in view of the uniformity of the vertebrate eye and the correspondences which have been found along other dimensions (**2**).

The second possibility is suggested by the photopic luminosity curve derived from the human observer (**5**). This curve indicates that the luminosity changes over the region of the spectrum from approximately 500 to 700 mμ may have distorted the gradients due to the influence of stimulus intensity dynamism. Since the data obtained from human *S*s exhibit the predicted relationship between discriminability and generalization, the phenomenon of luminosity need not be invoked to account for the results. For the pigeons, on the other hand, the gradients generated around CS values of 580 and 600 mμ should have shown asymmetry in the direction of decreasing wave lengths as a function of luminosity changes. Figure 1 indicates that the findings in the present study are contrary to the predictions made from luminosity changes over the spectrum. Moreover, a study by Blough[3] in which brightness differences were equated, revealed some changes in the form of the gradients when compared to those obtained by Guttman and Kalish (**7**, Fig. 1), but these changes were unrelated to predictions made from the discriminability function.

At the present time there does not appear to be any satisfactory explanation for the differences obtained between pigeon and human *S*s. The present study does suggest, however, that these differences may be related to the procedures used to obtain generalization (presentation of a single stimulus with the comparison stimulus absent) in contrast to those used to generate the discriminability function. The differences in generalization may further reflect differing characteristics of the species under these specific conditions.

[3] Donald Blough. Personal communication, March, 1958.

Summary

A previous study with pigeons, designed to explore the relationship between discriminability and generalization along the wave-length continuum suggested a lack of correspondence between these two processes. The present study is an attempt to extend these findings to human *Ss*. Forty *Ss* were required to judge the similarity among a series of wave lengths after presentation of the standard stimulus (either 530, 550, 560, or 580 mμ) for a period of 1 min. The judgment technique was the method of single stimuli with the comparison stimulus absent. Gradients of generalization around the standard stimulus were obtained. The shape of the gradients conformed to predictions made from the discriminability function for human *Ss* and indicated that the results from pigeons could not be generalized to human *Ss*. Moreover, the inverse relationship between generalization and discriminability was confirmed.

Several possibilities concerning the differences between predicted and obtained generalization gradients for humans and animals were discussed.

References

1. Bevan, W., & Saugstad, P. Breadth of experience, ease of discrimination and efficiency of generalization. *Brit. J. Psychol.*, 1955, **66,** 13–19.
2. Blough, D. S. Dark adaptation in the pigeon. *J. comp. physiol. Psychol.*, 1956, **49,** 425–430.
3. Brown, J. S., Bilodeau, E. A., & Baron, M. R. Bidirectional gradients in the strength of a generalized voluntary response to stimuli on a visual-spatial dimension. *J. exp. Psychol.*, 1951, **41,** 52–61.
4. Farnsworth, D. *The Farnsworth Dichotomous Test for Color Blindness*, Panel D-15. New York: Psychol. Corp.
5. Geldard, F. A. *The human senses.* New York: Wiley, 1953.
6. Guilford, J. P. *Psychometric methods.* New York: McGraw-Hill, 1936.
7. Guttman, N., & Kalish, H. I. Discriminability and stimulus generalization. *J. exp. Psychol.*, 1956, **51,** 79–88.
8. Guttman, N., & Kalish, H. I. Experiments in discrimination. *Sci. Amer.*, 1958, **198,** 77–82.
9. Hamilton, W. F., & Coleman, T. B. Trichromatic vision in the pigeon as illustrated by the spectral hue discrimination curve. *J. comp. Psychol.*, 1933, **15,** 183–191.
10. Hull, C. L. *Principles of behavior.* New York: Appleton-Century, 1943.
11. Lindquist, E. F. *Design and analysis of experiments in psychology and education.* New York: Houghton Mifflin, 1953.
12. Schlosberg, H., & Solomon, R. L. Latency of response in a choice discrimination. *J. exp. Psychol.*, 1943, **33,** 22–39.
13. Smith, K. Distribution-free statistical methods and the concept of power efficiency. In L. Festinger and D. Katz (Eds.), *Research methods in the behavioral sciences*: New York: Dryden Press, 1953.
14. Spence, K. W. The nature of discrimination learning in animals. *Psychol. Rev.*, 1936, **43,** 427–499.

(Received for early publication November 7, 1957)

JOURNAL OF EDUCATIONAL PSYCHOLOGY

This journal is now published by the American Psychological Association. In 1958 it became a bimonthly with issues in February, April, June, August, October, and December. Contents include articles on problems of teaching, learning, and the measurement of psychological development.

All back issues and subscriptions up to and including the May 1957 issue are the property of Warwick and York, Inc., 10 East Centre Street, Baltimore 2, Maryland.

Subscription $8.00
(Foreign, $8.50)

Single Copies, $1.50

Direct new subscriptions and renewals to:

AMERICAN PSYCHOLOGICAL ASSOCIATION
Publications Office
1333 Sixteenth Street, N. W.
Washington 6, D. C.